DOCUMENT RESUME

ED 434 130 TM 030 092

AUTHOR	Glimpse, Warren G.
TITLE	Decennial Census School District Project Planning Report. Working Paper Series.
INSTITUTION	Westat, Inc., Rockville, MD.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-WP-98-07
PUB DATE	1998-05-00
NOTE	60p.
AVAILABLE FROM	U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Avenue, N.W., Room 400, Washington, DC 20208-5652; Tel: 202-219-1831.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Census Figures; *Data Collection; Elementary Secondary Education; National Surveys; *Planning; *Research Methodology; Research Reports; *School Districts
IDENTIFIERS	National Center for Education Statistics

ABSTRACT

This document provides background information related to the 1990 Census Mapping Project to assist in evaluating the processes, products, users, and users of data from that program as part of the planning activity for the corresponding Census 2000 activity. Analysis of the data items frequently used, with particular attention to the School District Data Book, resulted in some key recommendations for a plan to develop a minimal 2001 census school districts special tabulation. The Census 2000 school district special tabulation should be used to produce several public use data products reflecting boundaries as of the school year 1999-2000. The public data base should be released on CD-ROM no later than December 23, 2001. It is also suggested that a summary report be prepared using these data to assess and report on the condition of education in the United States. An appendix lists the top 100 file subject matter items. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 434 130

NATIONAL CENTER FOR EDUCATION STATISTICS

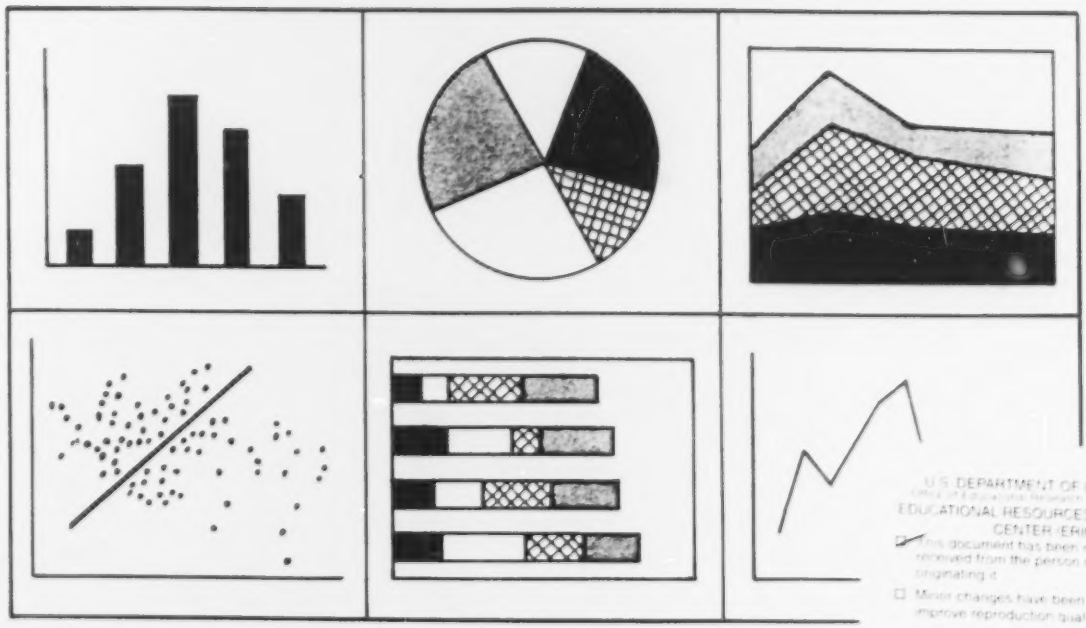
Working Paper Series

Decennial Census School District Project

Planning Report

Working Paper No. 98-07

May 1998



TM030092

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

U.S. Department of Education
Office of Educational Research and Improvement

BEST COPY AVAILABLE

Decennial Census School District Project

Planning Report

Working Paper No. 98-07 May 1998

Contact: Tai Phan
Surveys and Cooperative Systems Group
(202) 219-1627
e-mail: tai_phan@ed.gov

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Surveys and Cooperative Systems Group
Paul D. Planchon
Associate Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *Decennial Census School District Project Planning Report*, Working Paper No. 98-07, by Warren G. Glimpse. Project Officer, Tai Phan. Washington, D.C.: 1998.

May 1998

Foreword

Each year a large number of written documents are generated by NCES staff and individuals commissioned by NCES which provide preliminary analyses of survey results and address technical, methodological, and evaluation issues. Even though they are not formally published, these documents reflect a tremendous amount of unique expertise, knowledge, and experience.

The ***Working Paper Series*** was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series. Consequently, we encourage users of the series to consult the individual authors for citations.

To receive information about submitting manuscripts or obtaining copies of the series, please contact Ruth R. Harris at (202) 219-1831 or U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave., N.W., Room 400, Washington, D.C. 20208-5654.

Marilyn McMillen
Chief Statistician
Statistical Standards and Services Group

Samuel S. Peng
Director
Methodology, Training, and Service Program

Decennial Census School District Project

Planning Report

Prepared by:

Warren G. Glimpse
(under a consulting arrangement with Westat, Inc.)

Prepared for:

U.S. Department of Education
Office of Educational Research and Development
National Center for Education Statistics

May 1998

Acknowledgments

This document has been prepared by Warren G. Glimpse under a consulting arrangement with Westat pursuant to Westat's consulting services contract with the National Center for Education Statistics, U.S. Department of Education. For additional information regarding the Westat project activities, contact Ms. Pat Butler, Westat, 1650 Research Boulevard, Rockville, MD 20850 (301/251-1500). Warren Glimpse may be contacted at P.O. Box 3565, Alexandria, VA 22302-3565 (703/379-4700).

The review, comments and opinions of earlier versions of this document provided by Dan Kasprzyk, Kerry Gruber, and Tai Phan of the National Center for Education Statistics contributed greatly to the final version of this document. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

Table of Contents

Foreword iii
Summary of Key Recommendations ix

1. Introduction **1**

2. Background **2**
2.1. Authorization and Central Purpose 2
2.2. 1990 Census Mapping Project – the broader framework 2
2.2.1. Scope 2
2.2.2. School District Defined 3
2.3. School District-Level Statistics from the Decennial Census 4
2.3.1. State-by-State Detail Data Public Use Files 6
2.3.1.1. Types of Tabulation Records 6
2.3.1.2. Enrollment Iterations 6
2.3.1.3. Grade and Age Iterations 7
2.3.2. U.S. Summary ("Top 100") Public Use File 7

3. Subject Matter Evaluation **9**
3.1. User Diversity -- Requirements of different types of users 9
3.1.1. Usage by Type of Use 9
3.1.2. Usage by Type of User 9
3.2. Which Items are Used Most Often? 11
3.2.1. Items Used Most Often–Summary 11
3.2.2. Items Used Most Often–Tables Identified 11
3.2.2.1. Record Type 1–H -- Households 13
3.2.2.2. Record Type 2–PT -- Persons 13
3.2.2.3. Record Type 2–PS Persons 13
3.2.2.4. Record Type 3–HC -- Households with Children 13
3.2.2.5. Record Type 4–PR -- Parents with Children 14
3.2.2.6. Record Type 5–CH -- Children's Households 14
3.2.2.7. Record Type 6–CP -- Children's Parents 14
3.2.2.8. Record Type 7–CO -- Children's Own Characteristics 14
3.3. Which Items Are Used Least? 15
3.4. Which record types are used most often? 15
3.5. Accuracy of the Data 15
3.5.1. Universe tabulation specifications 16
3.5.2. Statistical accuracy 17
3.6. Issues concerning end use of the data 17
3.6.1. Evaluating level of use as descriptive, comparative versus analytical 17
3.6.2. Ease of applicability of traditional analytical methods 18
3.7. Geography Issues 19
3.7.1. Split Block Issue 19
3.7.1.1. Reasons split blocks should be used 19
3.7.1.2. Reasons split blocks should not be used 20
3.7.2. Block Cross Reference File 20
3.8. Linkage between 1990 and 2000 21
3.8.1. Maintenance of relevant children concept 21
3.9. Linkage with related subject matter 21
3.9.1. Common Core of Data 21
3.9.2. Census of Governments School District Finances 22

3.9.3. Other data sources 23
3.10. Annualized Updates 23
3.10.1. Demographics (key subset) 23
3.10.2. Boundary files 24
3.10.3. Access methods/linkages 24
3.10.4. Longitudinal analysis 25

4. Census Tabulation Universes and Iterations 26
4.1. Universes of tabulation 26
4.2. Relevant children 26
4.3. Enrollment Iterations 26
4.4. Age/Grade iterations 27
4.4.1. Grade Iterations 27
4.4.2. Age Iterations 27

5. File Structure and Media Alternatives 29
5.1. File Structure 29
5.1.1. Alternatives to file structure and table formats 29
5.1.1.1. Use of Data Compression 29
5.1.2. DBMS Alternatives 29
5.1.2.1. Selection of a database management system 30
5.1.2.2. Vendor Independent DBMS Options 30
5.1.3. Widely used transportable formats/structures 30
5.2. Media alternatives 31
5.2.1. Print-related media 31
5.2.1.1. Hardcopy 31
5.2.1.2. Print-oriented 31
5.2.2. CD-ROM 31
5.2.3. Internet 32

6. User Feedback 33
6.1. Most frequently asked questions on help/technical assistance line 33
6.1.1. Basic Questions About Existence of Specific Subject Matter 33
6.1.2. Issues Concerning Mapping and Graphics 35
6.1.3. Documentation 35
6.2. Most frequently reported problems and complaints 36
6.2.1. Use of CD-ROMs 36
6.2.2. Non-windows platform 37
6.2.3. Operation on Apple/Mac 37
6.2.4. Specific Data Desired Not Available 37
6.2.5. Desired Software Feature Not Available 37

7. Appendix -- Top 100 File Subject Matter Items 39

Summary of Key Recommendations

This page provides a summary of key recommendations. This page is not an executive summary of the report. Rather, it summarizes key points that are recommended for a plan to develop a minimal 2000 census school districts special tabulation.

1. A Census 2000 school district special tabulation should be prepared resulting in possibly several public use data products.

2. The resulting public use data products should be conveyed as information resources of the Department of Education and provided under the general name "Census 2000 School District Project."

3. Tabulations should reflect boundaries as of the school year 1999-2000.

4. Tabulations should make use of identical methodological procedures applied by Census for the 1990 Census school district special tabulation with regard to use of split blocks, definition of relevant children, grade augmentation and treatment of the PK and K enrollment attributes.

5. The public use database should be released minimally on CD-ROM medium making use of compressed ASCII records which can be uncompressed generating a CSV structured state by school district data file on the user's computer.

    - A preferred alternative or parallel public use product would make the master database available on Internet under database management enabling extract of the total database into multiple output formats.

    - A recommended alternative/supplementary form of access is through provision of stand-alone or networked computer system enabling longitudinal analysis (1990-2000 census data) linked with intercensal demographic updates, with capacity to link CCD and F-33 data.

6. The special tabulation database should contain, minimally, tables recommended in this document for "record types 5 through 7"

7. Public use data files should become available no later than December 31, 2001. It is recognized that the Census Bureau would traditionally make such data available at a later date due to various processing considerations. Prospective changing school district boundaries in some areas and sensitivity relating to single year of age specific cohorts make it important to have these data accessible as early as possible. Efforts should be made to preplan special tabulations through the use of the Census Bureau Data Access and Dissemination System and release data earlier than with previous decennial census special tabulation processing.

8. A summary report should be prepared using these data which assesses and reports on the condition of education in America. This report should be prepared by or commissioned by the Department of Education and made available to the public.

1. Introduction

This document has been prepared to provide background information relating to the 1990 Census Mapping Project to assist in evaluating the processes, products, uses and users of data from that program as a part of the planning activity for the corresponding Census 2000 activity.

Accordingly, this document is written for readers who are in some way stakeholders in this process. Since this is group of analysts, administrators and others who have varied experiences and familiarity with the background. A background summary of the 1990 Census Mapping Project is presented first. An understanding of many of the technical issues, definitions and processing of the 1990 census-related activity is essential to evaluating what should be done differently in 2000.

Primary topics to be addressed in this document have been set forth by the National Center for Education Statistics (NCES). NCES has asked that the paper address issues related to the content (data items) and the format (file structure) of the 1990 Census Mapping products, in particular, the School District Data Book (SDDB). The paper is also to address current technology applications that might make Census 2000 products more adaptable and user friendly.

Topics to be addressed in this document include:

- Which data items are used most often?
- Which record types are used most often?
- What are the most frequently asked questions on the help/technical assistance line?
- What are the most frequently reported problems or complaints?
- Is the SDDB subject matter too complex to be packaged and distributed on CD-ROM?
- What might be done differently in laying out file structure and table formats as the system is redesigned to accommodate user input and take advantage of emerging technologies?

Sections 1 and 2 deal primarily with what happened with regard to the 1990 census experience. Subsequent sections provide evaluations of these activities and developments in several contexts as set forth above.

2. Background

2.1. Authorization and Central Purpose

Authorization for the 1990 Census Mapping Project and development of the School District Data Book, as mandated in the scope of work description, was made in the following Public Law specifications.

"Among other purposes, these products are required for implementation of a Congressional mandate in Public Law 100-297 April 28, 1988, The Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvements Amendments of 1988, to wit:

> 'The Elementary and Secondary Education Act of 1965 (20 USC 2701 et seq.) ... is amended to read as follows:
>
> Sec. 3010) DECENNIAL ANALYSIS OF SCHOOL DISTRICTS.—Section 406(g) is amended by adding after paragraph (5) (as added by subsection (i)) the following new paragraph:
>
> > '(6) on April 1, 1993, and every 10 years thereafter, the Center shall submit a report to the appropriate committees of the Congress concerning the social and economic status of children who reside in the areas served by different local education agencies. Such report shall be based on data collected during the MOST recent decennial census."

The specific services required to facilitate the implementation of this mandate are to:" The text then proceeds to set forth tasks associated with development of the School District Data Book.

2.2. 1990 Census Mapping Project – the broader framework

2.2.1: Scope

The 1990 Census Mapping Project started in 1988. Under this initiative, sponsored by the National Center for Education Statistics and coordinated by the Council of Chief State School Officers, all states participated in a program to develop school district maps. The maps, the first complete set ever to be developed for the nation, were the critical first step in the development of the database.

As this project evolved, the named was retained – 1990 Census Mapping Project. Actually the scope of the project took on distinct but related parts:

- Mapping--initial mapping activity and development of the school district coding in the Census Bureau TIGER files
- Special tabulation--the process of developing the 1990 census school district special tabulation
- Public use products--development of the School District Data Book
- Analyses and user support

2.2.2. School District Defined

For the 1990 Census Mapping Project, a public school district is an area whose public schools are administratively affiliated with a local education agency recognized by the state education agency as responsible for implementing the state's elementary and secondary public education program. Through the Census Mapping Project, approximately 15,000 school districts were mapped.

School districts delineated by the 1990 Census Mapping Project are usually the same as those referenced as regular school districts in the NCES Common Core of Data Program. The Census Mapping Project used names and codes from the 1989-90 Common Core of Data as a means of identification.

Most areas of the U.S. are covered by one or more school districts. However, there are parts of some states that are not covered by any school district. These 60 areas are referred to as "balance of county" areas and treated as "pseudo" school districts in the SDDB. As a result, all areas of the U.S. are accounted for through the Census Mapping Project.

Paper maps developed by individual states were sent to the U.S. Bureau of the Census. The Census Bureau digitized the maps and transferred the resulting data into the Census Bureau's TIGER System. The TIGER (Topologically Integrated Geographic Encoding and Referencing) System is used by the Census Bureau as a way of tabulating address-oriented data. Once the school district maps were a part of the TIGER system, each of the nation's 6.5 million census blocks could be uniquely associated with their respective school districts and the census data could be assembled for all these districts.

Boundary definitions used for school districts in the School District Data Book originate with the Census Mapping Project beginning in 1988. Under this initiative, sponsored by the National Center for Education Statistics and coordinated by the Council of Chief State School Officers, all states participated in a program to develop school district maps. The maps, the first complete set ever to be developed for the nation, were the critical first step in the development of the database.

Pseudo Districts. Some geographic areas mapped for the 1990 Census Mapping Project do not technically correspond to true school districts. These areas are referred to as pseudo districts. In addition to the 14,985 school districts mapped, 30 pseudo districts, referred to as "balance-of-county" areas, were also mapped. Additionally, 268 pseudo districts corresponding to sub-district (areas which sub-divide true school districts) areas were mapped. Also, in California, the only state not to fully participate in the Census Mapping Project, 21 pseudo districts exist corresponding to county equivalent areas. Each of these types of pseudo districts are further described below.

Balance-of-County Pseudo Districts. Most areas of the U.S. are covered by one or more school districts. However, there are parts of some states that are not covered by any school district. These 30 areas are referred to as "balance-of-county" areas and treated as "pseudo" school districts in the SDDB. Balance-of-county areas, as the name suggests, are the residual part(s) of a county not assigned to any school district. Although these areas are treated as one area within a county for data tabulation purposes, in most cases a balance-of-county area is actually comprised of several non-contiguous parts.

While balance-of-county areas are not recognized by the State or Federal Government as true school districts, data have been tabulated for these areas in the 1990 census school district special tabulation. Due to the inclusion of these balance-of-county areas, all areas of the U.S. are accounted for through the Census Mapping Project and 1990 census school district tabulations. Balance-of-county pseudo districts have district codes that begin with the two characters 81 (no other district codes in the U.S. have 81 as the starting digits).

Sub-District Pseudo Districts. In Hawaii, which is collectively one school district, under the Census Mapping Project, 231 sub-districts were mapped. In New York, 37 sub-districts, within the New York City Public School District were mapped (32 community school districts and 5 borough secondary school districts). As both true

district and sub-district data are presented in the SDDB database, users may need to take special precautionary measures not to double count the data for these areas.

California Pseudo Districts. The State of California is the only state not to fully participate in the Census Mapping Project. As a result, some school districts in California were not mapped, nor were census data tabulated for these areas as true school districts. Twelve California counties, which are comprised of two or more true districts, were not mapped at the district level (Butte, El Dorado, Humboldt, Kings, Madera, Mariposa, Monterey, Napa, San Benito, Santa Barbara, Siskiyou, Tehama and Trinity). In these cases, the county itself was used as a pseudo district for census data tabulation purposes. Since data were tabulated for these county-wide pseudo district areas, the demographic data are available for California on a statewide basis even though data are not separately provided for many true districts. The number of true districts in California as conveyed by the Census Mapping Project is approximately 20-percent fewer than the true number of districts in the State.

Types of Districts. School districts are classified as elementary, secondary and consolidated (unified) based on the grade range administered by the district. Each school district's grade range in the 1989-90 Common Core of Data Public Education Agency master list represents the lowest and highest grades with non-zero student counts in the schools operated by the agency. Grades recognized for inclusion in the universe of elementary and secondary agencies range from prekindergarten (PK) through grade twelve (12).

The 15,006 school districts covered by the Census Mapping Project are categorized as follows:

- 11,269 consolidated districts
- 3,175 elementary districts
- 562 secondary districts

2.3. School District-Level Statistics from the Decennial Census

Pursuant to the census mapping project, school district level statistics form the 1990 decennial census were developed as a special tabulation--the 1990 census school district special tabulation. A special tabulation refers to a tabulation of data prepared using the basic record file of the decennial census. The basic record file contains individual respondent data. As these data are confidential, only the Census Bureau has access to the basic record file. Thus, only the Census Bureau can prepare "special tabulations."

The 1990 Census school district special tabulations are summary data. It contains only aggregate data describing attributes of groups of persons and households in school districts, counties, states and the nation. *There are no data about individuals in the special tabulation.*

The 1990 Census school district special tabulation is the largest special tabulation ever compiled by the Census Bureau. It is also the largest demographic database ever developed from a decennial census.

The two features of the 1990 Census school district special tabulation that set it apart from other special tabulations are (1) the way the data were estimated for school districts (hence the geography) and (2) the nature of the record structures and universe of tabulation (hence the subject matter specifications).

Geography. Only the 1990 Census school district special tabulation has used a process of splitting census blocks to develop estimates for the target area of tabulation (school district). In this regard it is unique both in complexity and methodology.

Subject Matter. Conventional Census tabulation records are developed for persons or households. In the 1990 Census school district special tabulation, the scope of person and household record types were extended. Of primary interest was where the "persons record" was developed specifically for children (persons 3-to-19 years of age and not a high school graduate). Only the 1990 Census school district special tabulation provides

detailed demographic data on children--for any point in time or for any type of geographic area below the national level.

The Census Bureau delivered two types of special tabulation data for this project: File A and File D. Data from File A were used primarily by the U.S. Department of Education and States for administrative purposes. Data from File D were the source of the data developed for general public use and placed on public uses CD-ROM's.

With the exception of two data items (total population and total housing units), the File D data are all sample-based estimates. That is, the data items are subject to sampling variability as well as others sources of estimation error.

Despite the existence of sampling and nonsampling error sources in the data, in general the data from File D were determined to be impressively accurate. This fact is evidenced in part by comparing the complete population item with the sample-based estimate of total population for a school district of interest. Indeed, the accuracy of the estimates draw into question the accuracy of the administratively reported data from other sources in some districts.

The File A data were used by the budget office at the Department of Education for purposes of administering the distribution of funds to states on the basis of population or other demographic data. The File A data are in the public domain and were been distributed by the Department of Education to individual state departments of education as a part of the review process in finalizing the 1990 census special tabulations.

It is noted that the File A data were re-processed by the Census Bureau, at the request of the Department of Education, after errors in geography were found and subsequently corrected. As a result, it is possible that, in a few cases, the total population data in the final File A tabulations will differ from the total population in the File D, School District Data Book, database. The number of districts affected is small and the size of the differences, when they do exist, is small. Additional re-tabulation of the data in a variation of File A have been prepared since completion of the initial, final File A to reflect changes made to some school district boundaries throughout the 1990s.

The Census Bureau had to develop the 200 mainframe magnetic tape reels containing the File D data several times. During this process, spanning a course of several months, magnetic tape files would be processed and validated, outside of Census, sometimes involving most of the tapes, before locating a systematic error in the tabulations. The Census Bureau would then re-run the tabulations, place the data on tape and the tapes which would then be re-processed by the contractor to develop public use files, each time returning the tapes containing the errors. There were at least six versions of the special tabulation tapes developed by Census in this manner.

This experience with the 1990 census is very important with regard to planning for the counterpart Census 2000 products for several reasons. First, the requirement to re-run the tabulations took time as did the transfer and limited scope validation. The re-processing of the data resulted in the public availability of the data being much later than scheduled. Secondly, the additional processing created unforeseen expenses at the Census Bureau to re-process the data and by others engaged in producing the final public use data. An effort should be made to expand on the scope of quality control checks for Census 2000 to minimize the possibility of re-work and undetected errors.

After completion of the File D data, it was determined that a key subject matter universe (one set of record type 2 discussed below covering approximately 3,000 subject matter items for all school districts and counties) had not been provided by the Census Bureau in the final File D tapes. These data had to be provided in a separate set of tapes.

Data from the File D magnetic tapes were structured into state-by-state detailed files and a U.S. by state, county and district summary file (referred to at the "top 100" file) for distribution on CD-ROM.

2.3.1. State-by-State Detail Data Public Use Files

The state-by-state detail data public use files were organized on a state-by-state basis and contained data summarized at the state, county and school district levels. Where possible the data were organized as a single file and placed onto a CD-ROM with adjacent states. In several instances, the data volume was so large that the file for a state had to be separated onto multiple CD-ROM's.

The detail data public use files provide potentially for seven types of tabulation records. Each of these types of tabulation records correspond to a specific type of population or housing universe. Three of the seven types of tabulation records–those with attributes of children, children's parents or children's households were iterated, repeated, for various types of enrollment and age-grade levels.

2.3.1.1. Types of Tabulation Records

The basic universe of tabulation is conveyed by the following 7 types of tabulation records. The number at the right indicates the number of cells that are contained in the particular tabulation record. The "electronic codebook" or table outlines and descriptions for each of these tabulation records is provided on each public use CD-ROM as a text file and in a query structure. Table details are not described here.

Type	Tabulation	Items
1	Characteristics of All Households	981
2	Characteristics of All Persons	5,688
3	Characteristics of Households with Children	808
4	Characteristics of Parents with Children	3,187
5	Children's Households Characteristics	808
6	Children's Parents Characteristics	2,813
7	Children's Own Characteristics	2,271

As an example of interpreting the above table, there are 2,271 data items about children themselves (Children's Own Characteristics) for each geographic area of tabulation.

Roughly 70 percent of the data items in each record correspond to the Census Bureau subject matter tables used in the 1990 Census Summary Tape File 3. The additional tables follow similar numbering/reference nomenclature but have been defined by NCES to serve more specific data uses; for example, dropout population and at-risk populations.

For record types 3 through 7, the basic tabulation universe is children 3 to 19 years of age who are not high school graduates. Tabulation categories are further detailed by type of enrollment:

If the size of the population cohort was too small to develop estimates for record types 3 through 7, no record was included for that geographic area. The criteria for a cohort being too small for the Census Bureau to generate the record was set by the Census Bureau should be examined carefully in planning for the Census 2000 special tabulations.

2.3.1.2. Enrollment Iterations

For record types 3 through 7 tabulations were made for *relevant children* (children ages 3-19 years, not high school graduate), a concept discussed later in more detail. Tabulations were developed for total relevant children and each of several types of iterations, reviewed next.

Five types of enrollment iterations were developed for each of record types 3 through 7. In actuality, the CD-ROM files distributed only contained data for the first three of these iterations. Data for the categories of enrolled in private school and not enrolled are derived by the user or derived automatically by the software provided with the CD-ROMs.

Total Enrolled & Not Enrolled
Total Enrolled (Public & Private)
Enrolled in Public School
Enrolled in Private School
Not Enrolled

This means that the 2,271 data items for children's own characteristics in a district are available for each of the enrollment categories. As shown above, each of the data items are equally available for those enrolled in public school as well as those enrolled in private school.

For each type of enrollment category, as applicable for a school district age/grade coverage, in record types 3 through 7, the data are further broken down by the following age/grade categories:

2.3.1.3. Grade and Age Iterations

By Grade:
Total Relevant
Grade 1- 4
Grade 5- 8
Grade 9-12

By Age:
Kindergarten
Pre-Kindergarten
Age 0- 2 years
Age 3- 4 years
Age 5-13 years
Age 14-17 years
Age 18-19 years
Age 3-19 years
Age 5-17 years

This means that the 2,271 data items for children's own characteristics in a district are available both for all children and for each of these age/grade categories. As shown above, each of the data items are equally available for those enrolled in grades 1-4, for example, in public school as well as those enrolled in grades 1-4 in private school.

2.3.2. U.S. Summary ("Top 100") Public Use File

The "Top 100" database was developed to provide a compact file of key data items to be provided on each CD-ROM for each district, county, state and the U.S. These data have been drawn mainly from the Census school district special tabulation. They include:

Persons by Sex, Race and Other Selected Attributes
Families/Households by Selected Attributes
Housing Units by Selected Attributes
Economic Characteristics (selected items)

Dropouts and At-Risk Children
Attributes of Children
    Total Children (selected items)
    Children Enrolled in School (selected items)
    Children Enrolled in Public School (selected items)
Selected Administrative Items (Common Core of Data)
Selected Financial Items (Census of Governments)

Due to its importance in planning, the list of subject matter items included in the top 100 file are shown in an appendix.

3. Subject Matter Evaluation

3.1. User Diversity -- Requirements of different types of users

Use and user diversity plays an important role in determining how the Census 2000 data should be developed. This section reviews use and user diversity with regard to type of use and type of user.

3.1.1. Usage by Type of Use

Federal program administration. A primary use of the data by the U.S. Department of Education has been to use the data in File A for administrative purposes to determine distribution of funds. This use of the special tabulation data made no use of the public use data products. In addition, revisions and updates were made to the File A data for which no corresponding revisions were made in the public use data products.

Report to Congress. Another important use of the data by the U.S. Department of Education was to prepare a report to Congress on the condition of education. Here, most of the data, though not on a district by district basis, play an important role in being able to report on particular population cohorts of interest. Yet, even though the specific district data were not of direct concern for such applications, the district level data had to exist in order to examine whether or not a district belonged to a class of interest in an analysis.

Analyses of all districts. Many organizations are interested in performing analyses similar to those contained in NCES reports. Analyses of this type, running across all districts (in a state, region or the nation) for some detailed attribute, has been a frequent type of application in academic research settings. It is one for which the type of 1990 Census special tabulation data organization was not well suited. An improved way to address this type of user need is to package the data on magnetic tape in a manner structured for conventional mainframe applications.

Profile and analysis for one area. Another type of use involves a structured display and analysis of many attributes of one district, a set of districts a county or state. In this type of use, it is often of equal interest to be able to determine how the attributes of one area compare to another district or another area. This was the primary target design for the CD-ROM database structure.

The planning process for the Census 2000 school district tabulations and products should carefully consider the foregoing range of data uses. In addition, planning should take into account the experiences of users as to user diversity, available technology and requirements for support.

3.1.2. Usage by Type of User

School system administrators and community leaders. NCES sought to design the public use data products in a manner for ease of use by school system administrators and community leaders. This group of users is quite diverse. This diversity requires a data access capacity supporting the needs of those who might have:

- limited computer-related resources or capacities,
- little time to learn features of operation,
- little in the way of local support or assistance and
- a strong interest in understanding the nature of the community served by the district and how it compares to others.

Insights into the needs of this type of users were gained from the training/workshop sessions conducted for educators/administrators on uses of national databases, where SDDB was part of the larger set, in a session sponsored by Iowa department of education and University of Iowa. Results of this program reflected that an awareness as to content and uses of the SDDB data could most effectively be developed by showing how the data could be integrated and made immediately relevant to their district.

Policy analysis users. The decennial demographic data are important in many types of policy related analyses conducted within the Department of Education and elsewhere. Case study type briefs were planned by NCES to show how the data could be applied to many existing policy analysis applications. To improve usability of the Census 2000 data, development of policy brief applications and demonstrations could help many prospective users with policy related analyses.

State and regional education agencies. Use by state education agencies was widely mixed. The variation from state-to-state seems to reflect the nature of leadership and staffing. While all state education agencies have a strong data coordination and collection operation, few of them are equipped to support general research and assist users of education-related data. Some sub-state area education agencies, viewing themselves as service arms for a set of school districts, produced reports and promoted use of the data.

Use by Congress. For Census 2000, it would be helpful to produce a report for Congress supplemented with the single analytically-oriented CD-ROM and/or potentially the same detailed material on Internet. This would improve ways that Congressional offices could make use of the 1990 census school district special tabulation data. Where possible, the supplementary data and narrative would address how the data can be useful in existing or pending policy-related issues, legislation and operations.

Academic research and services users. One of the largest user groups were academic researchers making use of the data for funded research. While these users are also quite diverse within the group (some applications on at-risk children, some applications on public versus private issues, etc.) there are certain generalizations that can be drawn that suggest changes for Census 2000. In general these users are well-equipped with both knowledge-ability on how to process the raw data and computer facilities to process the data. This user community will no doubt again desire the data on Internet and/or on magnetic tape with easy national scope data extraction capacities to move the data of interest into a structure for use with their preferred software. This set of users will find it highly desirable to access the 1990 and 2000 decennial data in the same manner even though the data structures might be separate.

Some university-based state data center operations, and other university data access service units, added the data to their collection and provided wide ranging access and support. As the data aged and were not supported with updates, software nor cross integration with related education data, demand slackened and university involvement in this area has declined. Unless there is a systemization to integration with other data and data updates, the same pattern is expected for the Census 2000 data.

Association users. Associations perceived by NCES as being heavily involved with K-12 education operations were automatically sent a copy of all products. Some associations made some use of the products while others never made use of the data. The only first-hand knowledge we have of these users comes from calls made with questions. Association users tend to be "issue" oriented where the issue is determined either by member requests or a current policy/legislation issue. Seldom, it seems are association uses of a general and continuing research nature. It appears, that while both American Association of School Administrators and the National School Boards Association had some interest in making the data available through some type of query or on-demand basis to their members, neither actually implemented such programs. As a result, it is difficult to predict the nature of demands of association users for the Census 2000 data.

Private sector users. Private sectors users have a mix of interests in using the data. Like other categories of users, these conclusions are based on experiences with users and not a survey. One type of use has been market analysis– companies selling products and services to schools, school districts of households by school district made some use the data to assess potential market size and character. Another private sector use has been the blending of data derived from the special tabulation with data or related services sold by the company.

3.2. Which Items are Used Most Often?

This section provides a brief summary of the types of items used most often. A specific list of tables used most often is then reviewed. The sub-section describing tables used most often might be useful in developing an initial set of table specifications for the 2000 census school district special tabulations.

3.2.1. Items Used Most Often–Summary

Based on the experience of several demographers and application specialists, an effort was made to determine the most widely used data items in advance. These items were included in the "top 100" file. The top 100 file is a shorthand name given to the U.S. summary file that contained approximately 100 subject matter items for each state, school district and county in the U.S.

The top 100 file was widely used because of its basic coverage of the most widely used items and ease of use to meet a wide range of needs. The selection of items included in the top 100 file met most users basic needs. It can be said that all of the items in the top 100 file comprised the items used most often.

What was not included that should have been included? Among the items used most often that were not included in the top 100 file were the number of children by single year of age starting at age 0 through age 19. Even though the tabulation was conceived to be focused on the number of children ages 3 to 19 years, a large number of users wanted corresponding data on the number of persons (total and relevant) by single years of age.

3.2.2. Items Used Most Often–Tables Identified

A specific list of tables used most often is reviewed in this subsection. The list of tables used most often might be useful in developing an initial set of table specifications for the 2000 census school district special tabulations.

Terminology. Following long-standing tradition and nomenclature, data fields in the special tabulation files are organized by tables of data item(s) also referred to as cells. A table may contain one item (one data field) or several hundred items (data fields). All items within a table are tabulations (summary statistics) based on the unique universe of tabulation and subject matter qualifications relative to that particular table.

It is suggested that the 2000 census school district special tabulation follow the protocol used in other census summary statistic files so that tables contained within the special tabulation are concordant with possible tables contained in standard products 2000 census public use summary statistic files.

Review of Record Types. The review of tables presented here is organized by "record type." Record types have been reviewed in general in a previous section. The record types are reviewed in more detail below.

Type		Tabulation	Items	Metadata File
H	1	Characteristics of All Households	981	90SDH.DBF
P	2	Characteristics of All Persons	5,688	
		PT tables	2,501	90SDPT.DBF
		PS (supplemental) tables	3,187	90SDPS.DBF
HC	3	Characteristics of Households with Children	808	90SDHC.DBF
PR	4	Characteristics of Parents with Children	3,187	90SDPR.DBF
CH	5	Children's Households Characteristics	808	90SDCH.DBF
CP	6	Children's Parents Characteristics	2,813	90SDCP.DBF
CO	7	Children's Own Characteristics	2,271	90SDCO.DBF

Record types are referred to numerically and by 2-character abbreviation. The record types are shown in the above list in the left-most column. Thus, "CP" refers to the children's parents record which is also referred to as record type 7.

According to the Census Bureau's present plans, school districts will be one of the standard geographic areas of tabulation. Under this plan, many of the tables containing the more highly used data in record types 1 and 2 would be provided as standard tabulations (specifically those which are a part of the standard Census Bureau 1990 Census STF-3 counterpart for Census 2000).

Similarly, according to present plans, there would be no data tabulated as standard decennial products corresponding to record types 3 through 7. Record types 3 through 7 have the series of iterative, sub-universe, records tabulated corresponding to enrollment, age, and grade categorizations.

The strategy in planning for inclusion of data corresponding to record types 1 and 2 would be to have as many of the necessary items/tables included as a part of the standard products as possible. In doing so, this would (1) reduce confusion with respect to how to access and merge data and (2) reduce costs of developing essentially a redundant set of data. However, it is strongly recommended that some total population and housing tabulations be carried forward in the special tabulation, anyway, as a cross-check for accuracy in the validation processing of the final special tabulation files.

In broad overview of the arrangement of the tables in the 1990 census special tabulation records, it is noted that 808 housing-related items (organized as tables) are the same in record types 1, 3 and 5. It is recommended that this type of concordance be maintained for the 2000 census tabulations.

Similarly, there are 3,187 population related items (organized as tables) which are the same in record types 2 and 4. Here again, it is recommended that this type of concordance be maintained for the 2000 census tabulations.

Specific data items/tables used most frequently are listed/reviewed in the context of the metadata for the database. It is considered infeasible and undesirable to add this volume of text detail to this report. Moreover, if only the items used most are listed, the process/listing (1) raises the question as to what was perceived not to be used as frequently and (2) limits the possibility for a collaborative assessment by readers of this document and suggestions. Stated differently, by including a listing of all tables presented from the 1990 Census special tabulation, and marking those used most frequently, should help analysts plan for items to be included for the Census 2000 special tabulation. The listing is provided in a series of metadata files described below.

Eight metadata/"codebook" files are used in the following discussion. These files are in dBASE structure and correspond to the record type summary listing shown above. The name of the metadata file is listed in the far right column. Readers may access these files by using the self-unarchiving file supplied with this document (90SD.EXE) or downloading the archive file from Internet at www.sunspace.com/90sd.exe.

Readers may browse these metadata files with any dBASE file structure handler program. The structure of the file is as follows:

Field	Field Name	Width	Use
1	TABLE	6	table number
2	TITLE	1	has value "1" if part of title–a table title line
3	ITEM	10	item code
4	TEXT	68	text of table/item description
5	SEGMENT	6	record type
6	STATUS	1	not used
7	SEQ	6	not used
8	SEL	1	Use: coded 1 on first table line if suggested as "used frequently"

Users who examine these files with dBASE-like commands might find the command ***browse fields sel,table,text*** a useful way to examine the file contents and view the indication of which tables are suggested due to frequent use.

In the review that follows, data which are summary statistics relating to substitutions and allocations should all be retained as applicable to tables identified to frequent use and retention.

3.2.2.1. Record Type 1–H -- Households

Based on frequency of use, the only item suggested from this record type is H001. All other items which have an STF3 counterpart should be available for all school districts. An effort should be made on behalf of NCES that the STF3 counterpart does include the corresponding STF3 items as applicable. Supplemental items which augment STF3 items by race/ethnic or related classification were not used frequently. It is suggested that these items not be included as 2000 census school district special tabulation items. It is suggested that for quality control purposes that the complete count number of housing units and the unweighted sample count of housing units be included in the special tabulation.

3.2.2.2. Record Type 2–PT -- Persons

Based on frequency of use, the only items suggested from this record type are those for which there was no corresponding table contained in STF3. There is a series of tables starting at table number 183 which should be retained. These tables are coded with a "1" in the dBASE selection field (SEL) in the first title record. All other items which have an STF3 counterpart should be available for all school districts as a part of the standard census products. An effort should be made on behalf of NCES that the STF3 counterpart does include the corresponding STF3 items as applicable. Supplemental items which augment STF3 items by race/ethnic or related classification were not used frequently. It is suggested that these items not be included as 2000 census school district special tabulation items.

3.2.2.3. Record Type 2–PS Persons

Based on frequency of use, the only items suggested from this record type are those for which there was no corresponding table contained in STF3. There is a series of tables starting at table number 300 which should be retained. These tables are coded with a "1" in the dBASE selection field (SEL) in the first title record. Many of these items relate to poverty status, persons at-risk, age-detail pertinent to analyzing these cohorts and characteristics of mothers. All other items which have an STF3 counterpart should be available for all school districts as a part of the standard census products. An effort should be made on behalf of NCES that the STF3 counterpart does include the corresponding STF3 items as applicable. Supplemental items which augment STF3 items by race/ethnic or related classification were not used frequently. It is suggested that these items not be included as 2000 census school district special tabulation items.

3.2.2.4. Record Type 3–HC -- Households with Children

Based on frequency of use, there are only a few "baseline" tables suggested for retention. These tables correspond to household related characteristics in other records. Addition or deletion of other tables should coordinated in this respect. These tables are coded with a "1" in the dBASE selection field (SEL) in the first title record.

Not included in the frequently used selection for this record type are a series of tables supplementing the STF3 tables that might merit retention. These tables are numbers P200 and above and may be reviewed by readers in the

metadata file. While these items were seemingly not used frequently, they clearly have a role in the analysis of households with children and might be further examined.

3.2.2.5. Record Type 4–PR -- Parents with Children

Based on frequency of use, there are only a few "baseline" tables suggested for retention. These tables correspond to person related characteristics in other records. Addition or deletion of other tables should be coordinated in this respect. These tables are coded with a "1" in the dBASE selection field (SEL) in the first title record.

Not included in the frequently used selection for this record type are a series of tables supplementing the STF3 tables that might merit retention. These tables are numbers P183 and above and may be reviewed in the metadata file. While these items were seemingly not used frequently, they clearly have a role in the analysis of parents with children and might be further examined. The P183 and higher numbered tables are the same subject matter tables as contained in the type 2–PS record.

3.2.2.6. Record Type 5–CH -- Children's Households

Many of the tables in this record have been widely used, owing to the unique nature of the tabulation universe and its general interest to analysts and researchers. Most tables are suggested for retention. These tables are coded with a "1" in the dBASE selection field (SEL) in the first title record.

As with other records, STF3 table shells that had been iterated for detailed race/ethnic categories and poverty incidence have been omitted as they have been used much less. Also as with other records, there are a series of non-STF3 tables that were specifically added to support analysis of children in the 5 to 17 years cohort and metropolitan/urbanized area classifications. Where these tables do not involve large-scale iterations based on race/ethnicity or poverty incidence, it is suggested that they be retained.

3.2.2.7. Record Type 6–CP -- Children's Parents

Many of the tables in this record have been widely used, owing to the unique nature of the tabulation universe and its general interest to analysts and researchers. Most tables are suggested for retention. These tables are coded with a "1" in the dBASE selection field (SEL) in the first title record.

As with other records, STF3 table shells that had been iterated for detailed race/ethnic categories and poverty incidence have been omitted as they have been used much less. Also as with other records, there are a series of non-STF3 tables that were specifically added to support analysis of children's parents and metropolitan/urbanized area classifications. Where these tables do not involve large-scale iterations based on race/ethnicity or poverty incidence, it is suggested that they be retained.

3.2.2.8. Record Type 7–CO -- Children's Own Characteristics

Many of the tables in this record have been widely used, owing to the unique nature of the tabulation universe and its general interest to analysts and researchers. Most tables are suggested for retention. These tables are coded with a "1" in the dBASE selection field (SEL) in the first title record.

As with other records, STF3 table shells that had been iterated for detailed race/ethnic categories and poverty incidence have been omitted as they have been used much less. Also as with other records, there are a series of non-STF3 tables that were specifically added to support analysis of children's own characteristics, metropolitan/urbanized area classifications and characteristics of living environment. Where these tables do not involve large-

scale iterations based on race/ethnicity or poverty incidence, it is suggested that they be retained. Some tables, such as P197, are indeed quite long and detailed and the relevance might easily escape someone reviewing the structures. In many cases, such as with P197, it might not be the intrinsic value of a set of cells or an iteration within a larger table that is of value. Rather, it might be the enabling ability for an analyst to aggregate over several cells to derive a subject matter summary of great interest/value.

3.3. Which Items Are Used Least?

With respect to table/item specifications, items used least are those that are not included in the suggested/frequently used categories reviewed in the previous section. The items used least fall generally into three categories.

Age/Grade Iterations. First the age/grade iteration records were seldom used. This was probably for two reasons: one, the complexity of the database did not make it "worthwhile" to examine these more detailed attributes. In addition many users did not care for the age break/categories that were used.

Type of Enrollment Iteration. Second, the dominant type of enrollment category used was public enrollment. Here again, there was little awareness on the part of the average use that attributes of those enrolled in private school could be analyzed on par with those enrolled in public school or not enrolled. Beyond this, there seemed to be little interest in examining the categories total, total enrolled, and not enrolled.

Race/Ethnicity Detail. In all record types, but predominately record type 2, there was much more race/ethnic breakout than typically desired. In all likelihood, a slight augmentation of the race/ethnicity given in STF 3 would be adequate.

3.4. Which record types are used most often?

The record types used most frequently were record types 1, 2, 5, 6, and 7. Record types 3 and 4 were used least frequently.

As reported in the previous section, for record types 3 through 7, the iteration record type used most often for enrollment was the iteration "enrolled in public school." While the iteration "enrolled in private school" was not as widely used as the enrolled in public school iteration, the reason is probably twofold. First, most state and local education agencies seemed more concerned about characteristics of those enrolled in public school rather than those enrolled or enrolled in private school. Secondly, there seemed to be less general awareness about the existence of the enrolled in private school iteration. These factors, and the increasing importance of private education, lobby for the retention of five types of enrollment iteration data.

For record types 3 through 7, the age/grade iterations were relatively little used.

3.5. Accuracy of the Data

At the beginning of this section, it needs to be clearly stated that the author believes the data to be generally quite good and accurate. But, at the same time, it is essential to note that there are many ways to be mislead by estimates in this database and that users are often too cavalier in their use of the data.

Accuracy of the data is ascertained in several ways. Users often do not question the accuracy of the data other than in generality and "in passing." The conclusion reigns that "these are the best data available." Of course, data that are considerably inaccurate can produce disaster with respect to statistical inference.

The first method involves data built into the database to help the user evaluate accuracy. The database includes the complete count total population, unweighted sample count of persons, complete count number of housing units and unweighted sample count of housing units. The complete count number can be compared to the corresponding value of the sample-based estimate giving a rough measure of whether the sample yielded an accurate estimate of the complete count value. One way to improve on this for Census 2000 is to include data complete count total relevant population, unweighted sample count of relevant persons, complete count number of relevant persons housing units and unweighted sample count of relevant persons housing units. These items were not included in the 1990 census special tabulation.

The second method that users can use to examine accuracy is to use the traditional method of estimating the size of the standard error of the estimate and evaluating this number in the context of a confidence interval. This method could also be improved with the Census 2000 database by making subject matter error adjustment data available in a form that applies directly to the special tabulation. For the 1990 Census school district special tabulation, subject matter estimation error rate adjustments were provided only at the state level. For users examining estimation error that varies that type of subject matter this required use of the state level rates rather than rates for school districts or sub-state regions. A related interpretative aid would be the optional automatic computation and display of the 95-percent confidence limits for an estimate.

The third method that can be used to determine accuracy of the data is to compare the estimated number of total relevant children to the number of students reported from administrative data in the Common Core of Data (CCD) program. This method yields a rough measure of accuracy as the data from CCD. It is a rough measure because (1) the data are developed through administrative reporting rather than self-enumeration and (2) the data apply to a different point in the year relative to the decennial data. Despite these considerations, the fact that the CCD data did come from a separate source made the CCD data a popular means of evaluating overall decennial data accuracy (and vice-versa).

The fourth method that can be used is the test of reasonableness. It was found that some isolated errors in the process of developing estimates of per capita income resulted in substantial errors than could not be associated with traditional sampling errors or related errors of estimation. While the user of the 1990 census special tabulation must invoke this test of reasonableness for some known estimates, it would be desirable to eliminate the problem of erroneous estimates in the Census 2000 tabulations due to inadequate quality control in the tabulation processing. This can be achieved by placing tighter quality control measures on the production of the estimates.

3.5.1. Universe tabulation specifications

There are at least two types of problems associated with the accuracy of the data relating to universes of tabulation. These problems arise due to the Census Bureau not providing certain summary levels of data, requiring that these data be derived from files provided by Census.

State summaries not provided by Census. For the 1990 census school district special tabulation, there were no estimates provided by the Census Bureau at the state level of geography for File D. These records were not prepared for the 1990 Census school district special tabulation in File D requiring that data in county records be aggregated to state level records. This process did work adequately, though adding to expense, difficulty and possibility of error, except for distributional statistics such as medians. Since medians cannot be computed directly from summary statistics, medians had to be estimated for these records using a special algorithm that might produce different results than had the Census Bureau produced median estimates directly from the basic record files.

Two Enrollment Iteration Records not Tabulated. Similar to the absence of state level summary records, data records were not tabulated for two enrollment iterations: (1) enrolled in private schools and (2) not enrolled. The values for data fields in these records were derived by subtraction using the enrollment iteration records: total, total enrolled and not enrolled. As with the state total records, this process did work adequately, though adding to expense, difficulty and possibility of error, except for distributional statistics such as medians. Since medians

cannot be computed directly from summary statistics, medians had to be estimated for these records using a special algorithm that might produce different results than had the Census Bureau produced median estimates directly from the basic record files.

In the final analysis, it might have been that both cost and time required were higher to prepare the state level records as a process following the special tabulation processing. In addition, the absence of the state totals resulted in an absence of the availability of state control totals for use in validating user applications involving aggregation of county level data. For Census 2000, state level estimates and values for all iterations should be produced by Census as a part of the tabulation process.

3.5.2. Statistical accuracy

There are only two complete count subject matter items included in the special tabulation–total population and total housing units. As a result, this database consists almost entirely of data items subject to sampling errors and related estimation and non-estimation errors (such as undercount, treatment of split block allocations, grade augmentation, and errors in special tabulation programming). A very large number of school districts are quite small. Some districts have less than 20 persons in the entire district. Sample based estimates in such small geographic areas are of no value. The user must take into account not only the size of the district but also the size of certain population cohorts. That is, the district might have 1,000 students enrolled but only 10 of, say, a certain race. As a result, the race-related estimates would be of no value even though estimates for the total student population might be quite accurate. Stated differently, the size of the standard error associated with the estimate could be so large in such instances that an accurate estimate could not be determined.

As with the use of virtually all small area sample based estimates from the decennial census, there is a clear tendency on the part of users of these data not to examine the statistical accuracy of the estimates. While this does not suggest that a change be made in the development of the estimates from Census 2000 per se, it does suggest that different methods of alerting the user to statistical reliability of estimates should be put into place as a software and documentation feature.

3.6. Issues concerning end use of the data

Accurately assessing how the data will be accessed, processed and used helps determine how the database should be structured. It also helps determine priorities for software, documentation and user support that should accompany the data..

3.6.1. Evaluating level of use as descriptive, comparative versus analytical

NCES originally anticipated that the 1990 census school district special tabulation would be used for the following types of applications.

1. By the Department of Education in developing a report to Congress with a CD-ROM insert (the original impetus for the "School District Analysis Book" CD-ROM),
2. By state and local education agencies who want to know more about their districts(s) and how they compare to others and
3. By researchers seeking to use this unique source of demographics to analyze a wide range of issues.

The intended focus of the software and packaging of the data was to make immediately useful to a user with quite limited familiarity with statistics, software and computer applications. It was thought that research and analysis within the Department and by researchers more in general could deal with data access issues without the assistance of specialized software, documentation and user support.

Through the actual experience of dissemination and use of the data products, it was found that educators, school systems administrators and community leaders made relatively little direct use of the data. In many instances it appeared that this set of prospective users were either unaware of the data and/or its potential value. In other instances, this set of prospective users looked to technical staff or consultants to process the data and provide reports or undertake analyses. In summary, there was relatively little demand on the part of these users for simply an easy-to-access, book-like information product. In planning for Census 2000 this suggests that while an easy-to-use product needs to exist, that the data should be packaged specifically for more "advanced" uses and users.

Original plans called for the release of the 1990 census school district special tabulation on magnetic tape. As costs increased, this plan was dropped as it was unclear what the level of demand would be and the product could be effectively disseminated. Each national scope copy of the database on magnetic tape would require at least 175 reels of tape. One copy of the original tapes furnished to NCES was provided to the National Archives and Records Service (NARS). It is believed that NARS has not completed the standard archiving for these files for several reasons. First, NARS sought to determine exactly how the data on the tapes varied from the data on the publicly released CD-ROM's. This would have required a tedious and time consuming documentation process for which NARS was not budgeted. Secondly, the magnetic tapes contained compressed files requiring specialized software to decompress into fixed-length field and fixed-length record structure. At that time NARS did not have the programming resources and budget to transform the tape files into the archival structure required by NARS.

NCES had also planned to distribute the database, as structured on CD-ROM, through the Federal depository library system. This plan was also abandoned due to costs of reproducing and distributing the CD-ROMs. In planning for Census 2000, these considerations suggest that a 2-to-3 CD-ROM reduced set of data would both be popular and provide a cost effective alternative to supporting user access via CD-ROM though Federal depository libraries, state data centers, and related intermediaries.

3.6.2. Ease of applicability of traditional analytical methods

In order to support traditional analytical methods, the product design was two-fold. First, the more rudimentary and comparative analysis features would be a part of the package. Second, that the needs of users with more sophisticated requirements would be met by empowering them to export data from the CD-ROM into any form that they wanted. This general framework should again be the approach that is taken with Census 2000 developments.

Some users suggested that the files be structured for use with SAS (the Statistical Analysis System from SAS Institute)–a widely used and powerful software package. There are several problems with this suggestion–whether it be SAS or a competitive product such as SPSS (Statistical Package for the Social Sciences). Most importantly, most school districts and sub-state education agencies do not have the SAS (or comparable) product, would not purchase it and could not make use of it. This seems to be a likely trend that will continue with respect to planning for Census 2000.

3.7. Geography Issues

3.7.1. Split Block Issue

Should split census blocks be used for school district tabulations for Census 2000?

For the 1990 census special tabulation approximately half of the nation's school districts made use of "split" census block geography. (Despite this incidence; many states are not affected at all because the district boundaries conform to whole counties, places or townships.) Split census blocks result from superimposing a map with boundaries of school districts onto a map with boundaries of census blocks. Wherever the school district "cuts through" a census block, 2 or more split blocks are generated. Yet, for Census Bureau statistical purposes, tabulations are still made at the elemental unit being a whole census block. Thus, to derive the summary statistics for a school district which contains split blocks, housing units are prorated into a particular school district based on the geographic portion of a census block that is included in that school district.

Clearly this creates the possibility for large scale errors. The statistical estimation assumptions, for which there is little or no evidence to support pro or con, are implicitly that the housing units are, on average, randomly distributed in the block affected. In reality, this is hard to believe.

Consider this example: a block which turns out to contain 1,500 people as the result of the actual census–much larger than might have been expected. It turns out that this is a new apartment building which is entirely located in that part of the block in district A. For census mapping purposes, the school district boundary between district A and B splits through the geographic middle of the block. Using the split block methodology from the 1990 census, 750 of the people (assuming equal distribution of persons per unit) would be assigned in error to district B–they should have all been in district A.

Why do split blocks exist? There are at least two reasons. First, school district boundaries often follow non-visible boundaries. The Census Bureau insists, where possible, that all geographic boundaries used in standard tabulation units (the block being the smallest) be visible. The reason is the alleged importance of enumerators in all cases being able to see where a boundary exists (this is only important when an enumerator is faced with the need to see such a boundary).

The second reason stems from the unlikely re-mapping of school district boundaries to conform to census block boundaries. To implement a program to have all school districts adjust to visible boundaries for Census 2000 seems more than unlikely–*though some states might desire to opt to do this*. It is unlikely because there are no staff to develop such boundaries without new costs being incurred–costs by the states, costs by the Census Bureau and costs by the Department of Education. In addition, it is not likely that changing school district boundaries in this manner would be politically viable at the local level.

3.7.1.1. Reasons split blocks should be used

1. The Census Bureau requires census block geography to conform to visible boundaries. Many school district boundaries do not conform to visible boundaries.

2. Aside from the previous reason, to convert existing census blocks into a new set of smaller area blocks would increase initial Census Bureau geographic processing costs and add to the enumeration and census taking costs.

3.7.1.2. Reasons split blocks should not be used

1. The cost, time and error-likelihood in developing school district tabulations (processing by the Census Bureau) would be reduced if split block geography processing were not required.

2. The costs to the U.S. Department of Education for the school district special tabulation processing would be reduced if split block geography were not required. This costs consideration addresses only the costs of processing the special tabulation–not possible costs borne by the Department to develop new maps. The Department would not have to pay the Census Bureau to develop the block allocation processing and "conventional" special tabulation software, thus reducing special programming time, and specialized processing steps.

3. Boundaries for small area statistical blocks used for Census tabulations cannot be easily drawn.

4. Statistics tabulated for the split portions of the block have some unknown error rate (refer to example reviewed above).

5. Split blocks create the need for the Census Bureau to handle development of school district tabulations in a manner different that for other types of geographic areas which conform to whole block boundaries.

3.7.2. Block Cross Reference File

A cross reference file was needed by Census to produce tabulations for each school district. This file provided the block or split-block–smallest area geographic units–comprising the school district. This file has many uses that support and go beyond the data in the school district special tabulations per se and should be part of the Census 2000 public use products.

As an example of the importance of this file, consider the school district which needs to re-district its attendance areas in the year 2005. What the district needs to know is how many people lived in each of the smallest elemental units making up the school district. The reason for this might be that the attendance area boundaries need to be redrawn to conform to township (or some other) boundary and provide equal population representation based on Census 2000 data. That type of application is common and cannot be supported form the standard 1990 census public use products.

The block cross reference file should be patterned after a possible parallel public use master area reference file that will likely be developed by the Census Bureau for Census 2000. The block cross reference file should include all of the same code hierarchy as contained in the Census master area reference file. More specifically, all of the higher level codes corresponding to a block/split block, such as state, county, township, place, urbanized area, tract etc. should be carried in record as well as school district code. If this code hierarchy is omitted or partial it may require most users of the file to preprocess it before it is useful. Such preprocessing can be expensive and time consuming. For example, inserting the township code in the block/split block record after the file has been released without it would require processing another file that contained the tract/bna code and the block code. Also, it remains that most state use a school district code that differs from the Federal code. It would be useful if the state code is carried for the district is also included in the block/split block record.

3.8. Linkage between 1990 and 2000

There are many issues associated with linking 1990 and 2000 decennial data. Insuring that this longitudinal comparability exists is one of the most important features to augmentations of Census 2000 relative to the 1990 census school district special tabulations. If certain steps are taken, it will be possible to examine, for example, the dimensions of regional change that have taken place over the decade. For the first time it will be possible to undertake analyses such as:

- researching issues such as where sweeping changes are taking place in the school district population composition
- assessing how changes in the student population are taking place in terms of demographics (ESL requirements by language, at-risk children, poverty-related considerations, etc.)
- determining how long term changes in demographics at the district level parallel changes in achievement, economic development and related matters.

With exception of those states where school district boundaries coincide with place or county boundaries, it is not possible to compare the 1990 and 2000 decennial school district data with decennial data for previous censuses. This situation is largely the result of the way that school district boundaries and tabulations were developed before the 1990 census. That is, for historical tabulation processing reasons, it is not feasible to develop national scope school district data from decennial censuses prior to the 1990 census. In addition, the concept of relevant children, new to the 1990 census, creates a situation of double counting children who lived in areas covered by both elementary and secondary districts. Finally, the 1990 census marked the first time that tabulations of children's own characteristics were available.

3.8.1. Maintenance of relevant children concept

The tabulations of relevant children defined as 3 to 19 years of age and not high school graduate should be maintained to retain basic subject matter comparability. Relevant children should be defined as the grade relevant children for which as school district is responsible and the algorithm used by the Census Bureau to generate grade relevancy should be retained and applied in an identical manner.

3.9. Linkage with related subject matter

3.9.1. Common Core of Data

Administrative data from the 1989-90 Common Core of Data were included as separate files and included on each CD-ROM. A few items believed to be the more widely used, were added to the "top 100" file so that the items were in the same record for the school district. This made for a handy way to make use of a standardized, documented file with commonly available spreadsheet and database management programs.

Readers familiar with the NCES Common Core of Data may be aware that data from the basic file are released in a public use form annually. The SDDB CCD file presents similar subject matter data as the conventional annual file but the data have been organized differently.

This file was developed by the Census Bureau at the request of NCES using the CCD school level file. Using the school level data, school district level aggregates were prepared for various cross-categories involving the number of students by type, teachers by type and schools by type. This specialized version of the file was developed from the school level CCD file so that the number of schools could be tabulated by various criteria. For example, the

tabulation produces the count of schools that have fewer than 100 students, the number of schools that have between 100 and 1000 students and other such school-level cross tabulations. This file was developed because the CCD agency public use file does not contain these types of summary statistics.

The primary feature of these data of value and interest to users of the demographic data is that they provide data on the number of teachers, number of schools and number of students enrolled by school district and related attributes of students participating in Federally sponsored programs. The special aggregate developed from the school level file added value by including data on the number of schools by student size class and similar cross cuts.

School districts included in the 1990 Census school district special tabulation are not identical to the set of districts included in CCD. Initially, the entire CCD special aggregate file was included with the CCD. User confusion resulting from the existence of a district as a CCD district but not as a CMP district resulted in NCES requesting that only those CCD district records be included where there was a corresponding 1990 Census school district special tabulation district.

In planning for the Census 2000 school district products there should be access to a time series structured database containing the CCD data–as released in terms of the standard CCD districts file and the custom constructed aggregates developed from the school level file. That is, either the software should be available to construct a time series; for example, total number of teachers by school size class annually from 1995 through 2000, by accessing separate annually-oriented data files or a file should exist that has fields of annual number of teachers by school size class by district. There should be data for the 1989-90 school time dimension as closely defined as possible to the final set of districts for which data were tabulated for the 1990 Census special tabulation. Similarly, there should be data for the 1999-2000 school year constructed in a similar manner. These data should then be updated and added to the time series database on an annualized basis.

3.9.2. Census of Governments School District Finances

School district level financial data, from the 1989-90 Census of Governments Survey of School District Finances (a census in the 1989-90 school year) conducted by the Census Bureau, was included on the SDDB CD-ROM's. These data are often referred to by the short-hand reference F-33 which is the number assigned to the form used to collect these data.

Like the CCD data, a database file was included on each CD-ROM containing the all-U.S. F-33 data for all school districts. In addition, a few items were extracted and included in the "top 100 file.

The primary feature of these data of value and interest to users of the demographic data is that they provide data on the sources and uses of funds by school district.

Like the CCD data, school districts included in the 1990 Census school district special tabulation are not identical to the set of districts included in F-33. As with the CCD file, NCES requested that only those F-33 district records be included where there was a corresponding 1990 Census school district. In addition, there are some data anomalies with the F-33 data relative to the 1990 Census school district special tabulation data relating to the definition of some districts. Some community college districts (not a part of the 1990 Census school districts) included revenues and expenditures relating to 1990 Census school districts (not covered by F-33)

In planning for the Census 2000 school district special tabulation products there should be access to a times series structured database containing the F-33 data. That is, either the software should be available to construct a time series; for example, total expenditures annually from 1995 through 2000, by accessing separate annually-oriented data files or a file should exist that has fields of annual total expenditures by district. There should be data for the 1989-90 year as closely defined as possible to the final set of districts for which data were tabulated for the 1990 Census special tabulation. Similarly, there should be data for the 1999-2000 school year constructed in a similar manner. These data should then be updated and added to the time series database on an annualized basis.

3.9.3. Other data sources

While the CCD and F-33 data are some of the more popular data to use with the decennial/demographic data for analysis of schools and school districts, there are many other data sources popular for the same types of analysis. The capability for users to easily link these disparate data sources should be further examined.

There are two broad categories of data that should be examined as to how they might better be accessed and used with the decennial/demographics, CCD and F-33. First, school systems and State education agencies should be able to automatically interface up-to-date reporting data with these secondary data. Second there are a wide range of other survey based and administratively reported data maintained by NCES that should be interfaced with these data without having to extract and re-build data structures.

3.10. Annualized Updates

One of the most often requests of users of the demographic data from the school district special tabulation was for more up-to-date data and recent historical data enabling them to compare trends in recent years. Some users have no interest in using the data once they learn the data to be so out-of-date.

The Census Bureau is now in the process of developing a first set of school district intercensal demographics updates. The purpose of the school district intercensal estimates project is to develop statistics that might be used (at the option of the Secretary of the U.S. Department of Education) to allocate/distribute Title I funds to states and sub-state areas and administer other Federal programs. There is no explicit provision for public dissemination of these data. The scope of the subject matter is restricted to meet the narrow needs of the Federal programs involved and indeed have prospectively little usefulness in helping to assess the state of education at the district level nor facilitate planning for issues such as school improvement.

The Census Bureau plans to release the first set of school district intercensal estimates in the fall, 1998, to the Department of Education. These estimates will cover the following types of basic subject matter items:

- median household income
- number of people below the poverty level
- number of children under age 5 below the poverty level
- number of related children ages 5 to 17 years in families below the poverty level
- number of people under age 18 years below the poverty level

Despite its current scope of subject matter limitations, this program should be systematized and produce annual baseline updates. Starting in 2003 an annual program of district demographic updates should be implemented.

3.10.1. Demographics (key subset)

The demographic (population and housing) estimates now being prepared by the Census Bureau are being produced for possible use in distributing Title I funds and meeting related administrative requirements of the U.S. Department of Education. Others will want to use these data as more up-to-date estimates for non-administrative applications (e.g., size of at-risk population) and in longitudinal analyses (e.g., change in the number of relevant children in poverty). These data should be useful to help state and local education agencies monitor progress, change and assess new requirements. It is important to assess the scope of the basic subject matter items that should be included to meet to mix of needs that exist. Plans should be established soon for (1) when intercensal estimates will be prepared for school districts in the 2000's and (2) what methodologies will be used to prepare the estimates. While

state demographers develop intercensal estimates, none develop estimates for school districts (other than in states where school districts coincide with county boundaries).

3.10.2. Boundary files

The Census Bureau is new developing intercensal boundary updates for school districts. This work is undertaken primarily to support (1) retabulations of 1990 census data to reflect the current boundary for Federal administrative uses (2) for use in developing the intercensal estimates mentioned above.

There is no plan to make updated boundary files available for public use. Like the intercensal demographic estimates, the boundary files should be made available for public use. In some states many of the boundary files developed from the 1990 census school district special tabulation have changed so dramatically that the boundaries have only value to historical analyses–they do not characterize the boundaries as they exist today.

Priority should be given to release of annual public use school district boundary files structured in a manner that most mapping programs can use. Now, the program plan calls for updates to boundary files every two years. In this way, state education agencies and the local education agencies can be empowered to use the statistical data in thematic mapping applications using up-to-date data and boundaries. Users can also use maps for orienteering purposes–performing applications such as seeing where a block group or census tract is located with respect to the district boundary or where a highway or other visible earth surface feature is located relative to the district boundary. Other geographic processing applications, important to the school district's fiscal management could be more feasibly undertaken–such as possibly reducing costs through new types of transportation optimizing processes and reduction or deferral of capital expenditures through redistricting and attendance area analyses.

As a technical issue, it should also be noted that creating boundary files from the Census Bureau Topologically Integrated Geographic Encoding and Referencing (TIGER) Line files is not a matter of converting file formats or translation. In most mapping and geographic information system software, the boundary files used to draw polygons (school districts), represent the process of chaining together attributes from the TIGER records which form the boundary. While the Census Bureau (or NCES) could implement a "canned" process to create school district boundary files in a generic format for public use, no capacity exists today (it does, however, exist at the Census Bureau for other types of geography such as states, counties and congressional districts).

3.10.3. Access methods/linkages

It was the original intent that the School District Data Book would not be a fixed, one-time resource. Rather, it was envisioned that the baseline 1990 census data would be interfaced with many other types of data resources. It has turned out that the "top 100" file has been tacked onto other data files released by NCES such as the annually released CCD CD-ROM. While this approach has succeeded in making a limited set of the census school district special tabulation data accessible in a fundamental way, many users are not equipped to make use of the data.

The collective data packaging and access framework should be reassessed in developing a plan for Census 2000. There should be a database structure that enables users to access 1990 census and 2000 census data as well as a range of time-series oriented data from annual administrative reporting programs at the school district level. The database structure should follow internal structural standards set up for the Census Bureau Data Access and Dissemination System (DADS). The recommendations is not for purposes of recommending DADS as the public access system for accessing these data but rather insuring compatibility between the complex data structures. A plan for a system featuring integrated access via Internal, minimally, would feature the ability to access and mix data from each of these sources for user specific time periods.

3.10.4. Longitudinal analysis

To equip users with the best selection of capabilities for longitudinal analysis between the 1990 and 2000 censuses, a key set of 2,000 to 3,000 items from the 1990 census and 2000 census as the school district was defined at that point in time. The processing to tabulate the data items from Census 2000 should be done in a manner that provides for maximum consistency between all data items between the two points in time. While this database might be elaborately structured under a database management systems and be distributed with software the data should be released on CD-ROM and possibly other media in an easy-to-read and convert ASCII format/file structure.

Corresponding boundary files should be released which depict the boundary of the school district in the 1990 census and in the 2000 census. Like the demographic data, these boundary files should be released in a generic ASCII structure than can be imported by most mapping programs thereby providing users maximum flexibility with regard to use of the data and the boundary files.

To empower users who seek to analyze characteristics of districts which have changed geographically over the decade, the 1990 census and the 2000 census master area reference files should be prepared. These files would define the school district down to the split block level of geography. By using other Census Bureau summary data, it would be possible for users with sufficient resources to have a maximized capability to compare parts of changing school district areas over the decade. Thus, it would equip researchers with the capacity to track and assess demographic change by school district for user specified research applications.

4. Census Tabulation Universes and Iterations

4.1. Universes of tabulation

The universe of tabulation, described more generally earlier in the paper, are listed below.

Type	Tabulation	Items
1	Characteristics of All Households	981
2	Characteristics of All Persons	5,688
3	Characteristics of Households with Children	808
4	Characteristics of Parents with Children	3,187
5	Children's Households Characteristics	808
6	Children's Parents Characteristics	2,813
7	Children's Own Characteristics	2,271

Of these universe types, the most used were universe types 1, 2, 5, 6, and 7. The least used were universe types 3 and 4. Within the group of children's characteristics most use was made of record type 7, then 5, and then 6.

Rather than choosing which possible record types to eliminate, consideration should be given to dramatically reducing the number of cells associated with some of the record types. That is, data corresponding to record type 3 and 4 should not be eliminated but reduced 50-percent or more.

Item oriented selection issues and options are reviewed in a subsequent section.

4.2. Relevant children

Iterations for record type 3 through 7 are discussed in the remainder of this section. These are all iterations for the class of population referred to as ***relevant children***.

The definition of relevant children for the 2000 census should be maintained just as it was defined in the 1990. It must be defined as the number of persons in that district who are

1. within the age range of 3 to 19 years of age,
2. not high school graduates, and
3. within a grade for which that school district is responsible.

The methodology used by the Census Bureau to determine the grade in which the person should be within, should be the same grade augmentation methodology in the 2000 census as applied for the 1990 census special tabulation.

4.3. Enrollment Iterations

Total Enrolled & Not Enrolled
Total Enrolled (Public & Private)
Enrolled in Public School
Enrolled in Private School
Not Enrolled

Only the three iterations of total enrolled and not enrolled, total enrolled and public enrolled should be generated from the special tabulation (as was done with the 1990 census). The other two iterations, enrolled in private and not enrolled, are derivable–other than distributional statistics which must be estimated using a distributional estimation algorithm. *This is not a recommendation to do away with the enrolled in private school and not enrolled record iterations*. Rather, aside from control totals and distributional statistics, the field values would be derived from other iteration records as with the 1990 Census special tabulation.

4.4. Age/Grade iterations

For record types 3 through 7 age and grade iterations of the basic record of tabulation were generated (assuming there were sufficient observations to develop and estimate and confidentiality rules could be maintained).

4.4.1. Grade Iterations

Grade iterations are as follows:

Grade 1 - 4
Grade 5 - 8
Grade 9 -12

The grade iteration was of little value. It is recommended that this iteration be removed. The reason for arguing that a grade iteration should be maintained is only to support analysis of prospective restructuring issues with respect to elementary and secondary district configurations. No actual case of this is known to have taken place. If this is of interest, a better way would be to expand the age iteration records and eliminate the grade iterations records. It appears in retrospect that the grade iteration records were left in the tabulation because of interest voiced by factions who are no longer known today.

4.4.2. Age Iterations

Age iterations are as follows:

Kindergarten
Pre-Kindergarten
Age 0- 2 years
Age 3- 4 years
Age 5-13 years
Age 14-17 years
Age 18-19 years
Age 3-19 years
Age 5-17 years

For Departmental regulatory and administrative reasons, the iterations for 3-19 years and 5-17 years is necessary to be maintained. Moreover, since these categories relate to distribution of Federal funds by district, it is necessary to maintain the same categories for Census 2000 for longitudinal analysis and program evaluation.

Other age iterations are less important. Based on apparent categories of data use, the next two most important iterations would be for Kindergarten and Pre-Kindergarten.

For reasons of cost reduction and making the database more useable, it is recommended that five age iterations be eliminated. These iterations include:

Age 0- 2 years
Age 3- 4 years
Age 5-13 years
Age 14-17 years
Age 18-19 years

Note that the suggested removal of these iterations does not suggest that age-related detailed would not be included in the database. To the contrary, it is suggest that the age detail available in some tables be extended to single years of age.

Other than for the distributional statistics and control totals, the values could all be derived from the other combinations of age iterations. For the 1990 Census, these items were tabulated but not carried forward into the SDDB database. The values for these five iterations were derived by multi-record real-time processing. That is, when the user requested table(s) from one or more of these five iterations, the software retrieved the appropriate iteration records, performed the addition or subtraction real-time, while the user was waiting–instantaneous, and reported the derived values on the screen.

5. File Structure and Media Alternatives

5.1. File Structure

This section is mostly concerned with the data files from Census 2000 and less attention is given here to the matters of linking the CCD, F-33 and related data. These reason for this is that the dimension of the "problem" and issues that need to be addressed with first priority–in the context of this paper–relate to the Census 2000 file structural issues–accessing Census 2000 school district data.

5.1.1. Alternatives to file structure and table formats

In all likelihood, whatever, the level of detail, the Census 2000 school district tabulations will parallel those form the 1990 census in certain respects. That is, the basic subject matter fields will again parallel what was STF3 in the 1990 census. As a result, there will again be the structure, peculiar to the decennial census where there is a prevailing table/cell construct.

One of the most important features of data tagging, nomenclature and structural reference is that it parallels the standard Census 2000 summary statistic products. It would be inadvisable to introduce an incompatible file structure orientation where users have difficulty associating other decennial data.

5.1.1.1. Use of Data Compression

Data compression was applied to the 1990 census school district special tabulation resulting in a dramatic reduction in space requirement. Should the Census 2000 products make use of a data compression algorithm?

The answer as to whether data compression should or should not be used depends on two issues. First, will the size of the database be so large that there is no practical alternative? The answer to this question will be forthcoming once an initial set of tabulation specifications exists. Second, what will be the structure used for the Census 2000 summary statistic files and will that database architecture support the nature of universe iterations characteristic of the school district special tabulation.

5.1.2. DBMS Alternatives

Almost without question a sophisticated database management system should be used with Census 2000 school district database. The features of the DBMS should be feature relational database management capabilities and be widely used in the related subject matter fields. Complexity of the metadata for the Census-type data suggests that an important feature would be the scope and capacities that relate to metadata specifications and ease of user access. The selection of architecture should closely follow, and be consistent with, the developments used by the Census Bureau software being developed to support traditional summary statistics access. Unfortunately, the Census Bureau has not yet made any information available about these plans. This will be a development to watch and a close working relationship should be developed.

5.1.2.1. Selection of a database management system

There are many database management systems (DBMS) which might be selected for use with the Census 2000 data. The DBMS marketplace has grown dramatically throughout the 1990s. Issues surrounding this decision, and selection of a DBMS, should consider the Census Data Access and Dissemination System (DADS) developments, DBMS capacities supported within the Department of Education and ease of data transport to the user community preferred structures.

The following section relates to database structuring with respect to paralleling developments with the Census Bureau DADS. Census has yet to release an adequate description of the planned operating features of DADS, including:

- the ways in which the summary statistic files will be structured within DADS
- the operating platform for DADS
- the transportability of DADS for use on non-Census servers
- the nature of the metadata used to document the summary statistic files

These developments warrant a close watch by NCES on the DADS developments. To the extent possible NCES should have some representation on DADS development issues as some user support could be provided through DADS. Ideally, the school district special tabulation will be generated by DADS. The argument might be made by NCES to Census to prototype the DADS "special tabulation" capability using the school district special tabulation.

5.1.2.2. Vendor Independent DBMS Options

Failing a successful development with DADS, other options need to be examined. There are many application systems and languages which support vendor independent approaches to DBMS selection and structure. Use of a vendor independent approach to a data access system is highly desirable as it affords maximum flexibility to the largest number of users.

5.1.3. Widely used transportable formats/structures

With respect to dissemination of public use data files, this paper argues for the production of two types of data files/database structures. The master database would be developed for operation within the DBMS. Using the DBMS, a highly transportable version would be developed. Alternatively, the transportable version might be developable directly by the user community depending on the DBMS software operations.

Given that the Census DADS architecture seems viable the master database would be developed using structural specifications which is, or parallels, DADS. This might make access to the special tabulation data optionally supported by DADS by the general public. This matter will require continuous monitoring and development working with Census. If DADS proves not to be viable then an alternative DBMS structure should be developed. This might be done in coordination with a leading data warehousing system that possesses full-featured statistical query, analysis and display properties such as SAS from the SAS Institute.

In addition to having a DBMS and related access system there should also be provision to disseminate the more widely used data, the counterpart of the top 100 file or something more extensive, for use with widely available spreadsheet software and/or custom software. Probably the best way is to produce one comma-delimited file with associated tags and descriptors which can be imported by spreadsheet and similar software (CSV format). This strategy would assure the most widespread access to the most basic and widely used set of data.

5.2. Media alternatives

For the 1990 program, the only medium used was CD-ROM. For Census 2000, the types of media supported should be expanded.

5.2.1. Print-related media

This section reviews print-related media. Ideally, products reviewed in this section would be developed using the Census DADS, an augmentation of that system or a similar DBMS-related report generator.

5.2.1.1. Hardcopy

There should be one U.S. summary "statistical" printed report. The statistical report should be similar to those prepared by the Census Bureau but relate to a user audience focuses on children, their living environment and K-12 school systems.

Both of these reports should be in an Adobe PDF or counterpart structure which is available on Internet as well as conventional print. The hardcopy statistical report should have enough detail about each school district that key attributes can be easily looked up and viewed in a comparative manner. The statistical report to should have extensive graphics providing detail relating to the report to Congress but appeal to a wider set of user and serve as a single hardcopy reference document. Such a statistical summary report is badly needed.

This would also help enable the report to Congress focus more on the findings relating to a review of the "social and economic status of children who reside in areas served by local education agencies." Text in the report to Congress could be oriented to current and prospective Department of Education policies and programs and the implications for these programs' operations as suggested by demographic change and incidence of various key characteristics.

5.2.1.2. Print-oriented

There should be additional print-related reports which are actually Adobe PDF type files (or counterpart structure) prepared for each state. This document should be similar to the Census *Number of Inhabitants* reports and provide a basic set of numbers for each district. These reports would have more detail that the U.S. summary statistical report described above.

One of the items most in demand from the 1990 CMP was a concept the Census Bureau introduced, then withdrew, to develop "metafiles" containing predrawn maps showing school district boundaries co-existing with other types of geography. The term metafile should not be used as this term has now taken on more universal meaning with regard to a file describing data about data. The product would be similar to a PDF and enable the user to scroll across a "pre-printed" map for orienteering purposes.

5.2.2. CD-ROM

The question has arisen–is CD-ROM an appropriate media for use with the census school district special tabulations database? The answer is yes. The related question might be how should the data be differently structured on the CD-ROM. The main answer to this question has two parts. First, through attrition in the number of fields and record types that all data for a state fit onto a single CD-ROM. Second, a more sophisticated and easier to use data extraction tool be developed for use with the Census 2000 CD-ROMs. The data extraction software developed for use with the 1990 Census school district special tabulation is a good candidate for the Census 2000 data as well.

The access tool is fast, features maximum data record compression processing, has byte specific direct access features and can be interfaced into any type of Windows-based application. Other candidates for data extraction tools have similar properties with respect to the calling software expects the database structure to be in a certain form. The more transportable, and vendor independent DBMS compliant the database structure, the larger the database becomes–as a result of metadata and fixed field structure requirements.

5.2.3. Internet

The entire database should also be accessible via Internet server(s). There should be a plan for at least one central server operated by or for the U.S. Department of Education. There might be multiple mirror sites that would be operated at the expense of the hosting site. Ideally, these installations would be DADS or variations on DADS, given that DADS evolves into a viable system.

Internet would just be one more, albeit the most important one, networked based system for users to access the school district special tabulation system. That is, ideally, the structure of the host application system (DADS?) would be sufficiently transportable that it might be replicable on an Intranet. An Intranet capacity might be appealing to the Federal depository library as well as one operated by the OERI Labs. Reasons for this type of multi-co-located versions of the access systems are due to ways that the data might link with that network owner's own databases which make use of custom interfaces or proprietary data.

6. User Feedback

6.1. Most frequently asked questions on help/technical assistance line

This section describes some of the more frequently asked questions on the help/technical assistance line.

The more sophisticated and experienced users need the least help. The less experienced users may need a lot of help and not just in accessing the data, but in knowing about the significance of various parameters such as definitions, accuracy, proper usage, confidence intervals and other statistical matters of relevance.

In addition to this general characterization of users and need for help/assistance, it also varies by the nature of the application. For example, an application area of great importance is the subject area on school improvement and how these and other data may be used. There has become a major thrust in some states to develop indicators and related measures of student and educational success. These users will often be experts in their subject matter field and will be examining how to strategically make use of the decennial demographics and related data. Increasingly, the state agency is responsible for statewide indicators and then the local agency is responsible for the development and reporting of local district indicators. An assessment should be made of the state-by-state trends and how the state/local data collection/applications, such as the development of statewide education indicators reports, impinge on the design of examples of tutorials provided with the decennial census school district special tabulations database.

6.1.1. Basic Questions About Existence of Specific Subject Matter

Does this item exist? Perhaps the more frequent basic question has been–does a particular subject matter item exist in the database. If the answer is no, the follow on question is what data do exist in the database that might be a proxy or used as an indicator of the subject in question.

Despite the existence of a codebook and a very well developed index to subject matter tabulations, this facility, housed on the CD-ROM, was apparently used very little. In discussions with users inquiring about coverage of a particular item, the user would be directly to the CD-ROM self-help and more than half were not familiar with its existence.

How do I determine where to get this item? I need a measure of _____. If this does not exist, how can I approximate it?

I need the number of total relevant children by single years of age. I need the number of persons by single years of age in the district who are under 20 years of age. Questions of this type–for data on the number of persons by single year of age–was perhaps the most asked questions that could have been provided from the tabulation but was not provided.

Why aren't data here for my district? Names of districts often made it difficult for users not familiar with names, and certainly not the codes, to access data. Chicago, for example, was referenced as "City of Chicago." Some states make use of a name such as "District 35-J" which rendered the name unusable with regard to where it was located by community name or region.

In other instances districts (mainly in California) were simply not present in the database. This resulted from the fact that California did not fully participate in the census mapping project.

How can I get a current estimate for my district? Many users did not view this database as a "decennial product." They viewed it as the source of demographic data that they needed to prepare a grant, perform a needs assessment, analyze facilities requirements, etc. Their only interest was in getting (1) recent demographic estimates and (2) demographic data which matched the current boundary of the district.

How can I get names and addresses for schools? A remarkably large number of questions was received by users who wanted school level data. Often these questions were about subject matter, but most typically the interest was about the address or telephone number.

How can I get data for a metropolitan area? City? Urbanized area? Many users were unsatisfied with some aspect concerning the geographic level of tabulation. This request was typically of two types:

1. Many users wanted the data for metropolitan statistical areas, cities (comprised of multiple districts), congressional districts, area education agencies and urbanized areas. This class of users wanted data for larger geographic areas that what was provided. While some of these can be most generally aggregated for the data provided, others cannot (urbanized areas, congressional districts).

2. Other users wanted data for that part of the school district in a county or city, attendance area related summaries or parts by township. This class of users wanted data for smaller geographic areas.

How can I get demographics about teachers? There have sometimes been questions about specific data relating to demographics of teachers and staff. Interesting, there was a tabulation record planned on characteristics of teachers but this record was omitted during one of the initial re-runs by Census and was never added back into the process.

These data do not add up. Questions often involve clarification as to why data will not add-up and related topics. This topic is very important because there are many reasons for these questions and they will be one of the most repeated issue areas for Census 2000. Here are some examples of ways this question manifests itself.

1. The main source of this question results from mixing attributes of the whole population within a district with attributes of relevant persons.

2. The most typical question of this type was–when the population was summed for all districts in the state that the number of persons far exceeded the number of persons reported in Census reports for that state. This results from multiple districts mapping to the same area on the earth surface and is not an error, but is the source of confusion–it must be done the way that it was done. What needs improvement is ways to help users understand the definitions.

3. There is an error in the way the universe records were tabulated by Census (Under certain conditions data aggregated from the grade iterations for a district will not always equal total relevant children. This appears to have resulted for different ways the Census Bureau processed the grades PK and K in unified versus non-unified districts.).

I do not think these data are correct. There were many questions from users about the accuracy of the data. To be sure there are both isolated and systemic problems. But when most of these questions are assessed, it was found that there was a misunderstanding as to definition, the wrong geography or table had been retrieved, etc.

What are "at-risk" persons? While the question of definition about at-risk children was one of the most frequently asked, there were many others that followed this pattern of "basic questions." In large part these are the traditional types of questions asked by less experienced users about census data in general–what is a table, what is a household, etc.

6.1.2. Issues Concerning Mapping and Graphics

How can I print a map of my district? The basic mapping structure was designed to provide only a map of the state by district unless the user had substantial technical support. Many users desired to print a map of their district, ***easily***, in context of their region.

How can see a map of selected districts of interest? Related to the previous issue, many users wanted flexible mapping which was only available to users who both had the cartographic CD-ROMs, related full-featured mapping program and the time and technical support to use the supplemental products.

Digressing, the SDDB products were packaged and sold as:

1. the "statistical" CD-ROM's–44 CD-ROMs with access/display/extract software, and very basic mapping software, integrated on every CD-ROM,
2. the "cartographic" CD-ROM–7 CD-ROMs containing cartographic files with a file export program but no full featured mapping software and
3. the optional full-featured mapping software.

Thus, by design, to have full mapping features required the user to acquire the cartographic CD-ROMs and interface those files with their own mapping software or use mapping software provided for GIS applications.

How can I convert the boundary files for use with my software? Many users of these types of data already are familiar with a particular mapping program. In such instances, they desire to use the cartographic products, mainly school district boundary files, with their software. Even though an export routine was supplied to convert the boundary files into a more generic, transportable file, there is no truly generic structure. As a result, conversion often required specialized manipulation of the boundary files even after they were exported into the generic format.

How can I compare this district to other districts graphically? Many users asked to see added graphic display and analysis capabilities built right into the SDDB software. These capabilities were omitted with the thinking that many graphic display (bar charts, pie charts, etc.) already exist. Instead of offering this type of capacity, the plan called for the software to export a file from the database for loading into the preferred spreadsheet or make use of the "top 100" file which could be directly opened by all leading commercial software packages.

6.1.3. Documentation

NCES initially suggested than no document would be required. Our experiences suggest that the school district special tabulation requires extensive user documentation. Perhaps the ideal way to provide the documentation is with multiple layers, serving needs of different types of users rather than one massive reference document.

Basic documentation. A standard background summary with definitions and samples uses should be placed in the forward, body and appendices of the suggested printed reports described earlier in this report. This level of documentation would be fundamental, readily available and easily reproduced.

Level Two Documentation. A School District Data Book/1990 Census school district special tabulation user's guide/reference manual was produced in 1993, which had rather limited scope. A preferable approach to this "level two" documentation is to:

1. prepare a hardcopy manual, distributed not as an Adobe PDF, but as an ASCII or convertible file so that it can be used in a cut and paste manner with document prepared by the user, and

2. adapt this text into a help facility as used in a traditional Windows help facility.

Supplemental technical documentation. Some users have commented than no "electronic codebook" was available. In fact a very complete and sophisticated codebook was developed and was used as the core of the SDDB software. Its existence was and structure were not well documented though it resides in a freely accessible area on Internet and on every CD-ROM. This same type of detailed codebook should be available in a similarly structured manner for the census 2000 school district special tabulation.

Policy/analysis application briefs. A missing element to stimulate effective use of the 1990 census school district special tabulation were briefing type documents on how these data could be used to perform a wide variety of typically policy/analysis type applications. Prepared in the form of case studies, these short papers, also Internet accessible, could show how the data are acquired and used. In discussing the application, the user would be shown examples of how to use certain aspects of the data.

6.2. Most frequently reported problems and complaints

Despite the effort to develop an easy to use product, the most frequently reported problem was probably that the system/database was too difficult to use. While in part there are ways to improve this for Census 2000 with respect to software and user support mechanisms, it remains that we are trying to empower some of the most unexperienced data users to use one of the most largest and sophisticated mapping and demographic databases ever developed.

6.2.1. Use of CD-ROMs

I cannot get the CD-ROM to install. Many users experienced frustration with the install program. It could have been greatly improved but no upgrades were developed. The major source of frustration was that if the system did not install correctly the first time, directories were left on the computer which had to be manually removed before the install would work correctly. While this is a minor technical problem it proved to be a frustrating obstacle for less experienced or non-technically oriented users.

Do these operate on Apple/Mac? While MacIntosh computers can be equipped to run MS-DOS programs, the user experience was mixed with success. Also, not all version of Macs can run MS-DOS applications. Some require card add-ons which can make for cost, time and learning curve barriers for Mac users.

How do I extract data for the following items for all districts in my state? The data extraction feature provided to users did not fully meet their needs. The export capacity is an important weak point in the system and should have been improved with upgrades. The major problems included:

1. The requirements to develop the subject matter and geographic control files required for one type of data extraction were too much of a barrier for many users.

2. Many users were unaware that there was an automatic data extraction that did not require control file set-up. This feature went unused as readers tended not to read the documentation.

3. Many users wanted to set up a batch file and not make use of the interactive features of the system. This would be a relatively easy feature to add.

4. Having the data on CD-ROM required users to manually set up each CD-ROM for extraction. This set up requirement proved too technically challenging for many users or made it infeasible due to time, expense or simply frustration level.

6.2.2. Non-windows platform

NCES felt that the SDDB software not be developed using Windows-based methodologies. DOS applications were equally prevalent to Windows applications in 1993, and trends were unclear as to the future dominance of a particular operating system. The dominant view was that even if the computer was Windows-based, that the user could still make use of DOS-based application.

With the release of Pentium II computers operating with Windows Version 95 B, new problems have evolved. This version of Windows does not support execution of all DOS-based applications as did Windows 95 A. As a result, even though the SDDB software to use the CD-ROM continues to operate with Windows 95 Version B, the install program for the SDDB setup does not operate with Windows 95 B. The system must be "manually" installed using system commands or ported from another installation of SDDB available on another computer.

These developments are all common life-cycle issues. The correct plan for supporting the SDDB follow-on with Census 2000 is not to try to figure out what the best operating system will be but rather choose a development platform that is most likely to be able to be modified for different operating systems and platforms.

6.2.3. Operation on Apple/Mac

When the data were initially distributed, many users in education agencies and colleges and universities wanted to use the data on Apple/Mac computers. Many did not have access to IBM/PC computers. This remains an issue but the incidence has decreased since initial release of the data. Due to the expense of supporting an Apple/Mac platform , the increased availability of "WinTel" PC's, and the prospects for increased Internet access, this should not be considered an important priority for Census 2000.

6.2.4. Specific Data Desired Not Available

Subject matter items were often requested which fell outside the scope of the SDDB. These data involved topics such as facilities, staff attributes, achievement-related measures, etc.

Some suggestions are made in the listing of the top 100 file (see appendix) as to changes that might improve this file or clarify why certain data were provided as they were.

6.2.5. Desired Software Feature Not Available

Features of the software that was requested and not available include the following:

1. Sorting a list of district based on selected criteria.
2. Extracting data without having to create geographic and subject matter control files.
3. Extracting data based on a user specified criteria. For example, show me all districts in this state with percent of students enrolled in private school greater than 17percent.
4. Extracting district based on the county of location. Even though school districts do not conform to county boundaries in most states, many users would have found it desirable to select school districts for extraction is any part of the district was contained in a list of counties.
5. Aggregating districts selected to display a profile of aggregates.

6. Perform basic statistical operations on a set of selected districts (counties)

7. Process data on multiple CD-ROMs without having to manually set up a configuration file.

8. Preparing thematic maps in a more fully integrated manner. That is, suppose that the user selects 10 districts, knowing they are contiguous, the subject matter tables are then selected, the statistical data are then displayed. The user would like to click on a subject matter items and then an icon which instructs the software to draw a map of the selected 10 areas using the selected item to depict as a thematic pattern.

7. Appendix -- Top 100 File Subject Matter Items

The "Top 100" file should be expanded to include approximately another 100 items. The size of 100 items was arbitrary and experience suggests that a few more items would make this file much more useful.

This file included a record for each state, county and school district in the nation. Structured as a dBASE file for the 1990 census school district special tabulation, this file should be offered for the Census 2000 project as an integrated 1990 and 2000 census file. The file should be provided in (1) a delimited format for loading into most any program and (2) in Excel format with corresponding macros to perform query based structured profile displays

The numbers to the far right on each line show where the item referenced was derived from the master database – table number, record type and age/grade iteration.

Occupied Housing Units	H004	10 00
Vacant Housing Units	H004	10 00
Occupied Housing Units		
Urban - Inside Urbanized Area	H005	10 00
Urban - Outside Urbanized Area	H005	10 00
Rural - Farm	H005	10 00
Rural - Nonfarm	H005	10 00
Inside Metro - In Central City	H006	10 00
Inside Metro - Not in Central City - Urban	H006	10 00
Inside Metro - Not in Central City - Rural	H006	10 00
Outside Metro - Urban	H006	10 00
Outside Metro - Rural	H006	10 00
Owner Occupied	H008	10 00
Renter Occupied	H008	10 00
Median Gross Rent	H043A	10 00
Median Value	H061A	10 00
Families	P004	10 00
Households	P005	10 00
Median Household Income	P080A	10 00
Households		
with Public Assistance Income in 1989	P095	10 00
without Public Assistance Income in 1989	P095	10 00
with Children Under 18 Years	P200	10 00
with Children 5 to 17 Years	P201	10 00
Unweighted Sample Count of Housing Units	H002	1A 00
100-Percent Count of Housing Units	H003	1A 00
Unweighted Sample Count of Persons	P002	2A 00
100-Percent Count of Persons	P003	2A 00
Dropouts 16-19 NEIS & NHG - In Households	P061	2A 00
Dropouts 16-19 NEIS & NHG - In Group Quarters	P061	2A 00
Per Capita Income in 1989	P114A	2A 00
At Risk Pre-School Age Children - Less than 4 years	P300	2A 00
At Risk Pre-School Age Children - 4 to 5 years of age	P300	2A 00
At Risk School Age Children (6-19 years)	P304	2A 00
Persons with Income in 1989 Above Poverty Level	P117	2A 00
Persons with Income in 1989 Below Poverty Level	P117	2A 00
Total Persons	P001	2B 00
Urban - Inside Urbanized Area	P006	2B 00
Urban - Outside Urbanized Area	P006	2B 00
Rural - Farm	P006	2B 00
Rural - Nonfarm	P006	2B 00
Male	P007	2B 00
Female	P007	2B 00
NonHispanic White	P012	2B 00
NonHispanic Black	P012	2B 00
NonHispanic American Indian, Eskimo, Aleut	P012	2B 00
NonHispanic Asian and Pacific Islander	P012	2B 00
NonHispanic Other Races	P012	2B 00
Hispanic	P012	2B 00
in Group Quarters	P040	2B 00
16 Years and Over - In Labor Force	P070	2B 00
16 Years and Over - Civilian Employed	P070	2B 00
16 Years and Over - Civilian Unemployed	P070	2B 00

20 Years and Over by Educational Attainment		
12th Grade or less, no diploma	P188	2B 00
High school graduate	P188	2B 00
Some college, no bachelor or higher degree	P188	2B 00
Bachelor's or higher degree	P188	2B 00
Households with Relevant Children	P005	30 1F
Persons (Parents Living with Relevant Children)	P001	40 1F
Relevant Children		
Total	P001	70 1F
Urban - Inside Urbanized Area	P006	70 1F
Urban - Outside Urbanized Area	P006	70 1F
Rural - Farm	P006	70 1F
Rural - Nonfarm	P006	70 1F
Male	P007	70 1F
Female	P007	70 1F
NonHispanic White	P012	70 1F
NonHispanic Black	P012	70 1F
NonHispanic American Indian, Eskimo, Aleut	P012	70 1F
NonHispanic Asian and Pacific Islander	P012	70 1F
NonHispanic Other Races	P012	70 1F
Hispanic	P012	70 1F
Age 3 Years	P013A	70 1F
Age 4 Years	P013A	70 1F
Age 5 Years	P013A	70 1F
Ages 5-13 Years	P013A	70 1F
Ages 14-17 Years	P013A	70 1F
Ages 18-19 Years	P013A	70 1F
in Family Households		
Householder, Spouse, Grandchild, Other Relative, NonRel.	P017	70 1F
Child (natural, adopted, step)	P017	70 1F
in Non-Family Households	P017	70 1F
in Group Quarters	P017	70 1F
by Poverty Status - Income Above Poverty Level	P118	70 1F
by Poverty Status - Income Below Poverty Level	P118	70 1F
Ages 14-17 in Households	P017	70 14
Ages 14-17 in Group Quarters	P017	70 14
Children Enrolled in School		
Male	P007	70 2F
Female	P007	70 2F
NonHispanic White	P012	70 2F
NonHispanic Black	P012	70 2F
NonHispanic American Indian, Eskimo, Aleut	P012	70 2F
nonHispanic Asian and Pacific Islander	P012	70 2F
NonHispanic Other Races	P012	70 2F
Hispanic	P012	70 2F
Enrolled in Public School		
Male	P007	70 3F
Female	P007	70 3F
NonHispanic White	P012	70 3F
NonHispanic Black	P012	70 3F
NonHispanic American Indian, Eskimo, Aleut	P012	70 3F
NonHispanic Asian and Pacific Islander	P012	70 3F
NonHispanic Other Races	P012	70 3F
Hispanic	P012	70 3F

CCD Membership from File A
CCD Students
CCD Teachers
CCD Schools

F-33 Total Revenue
F-33 Local Revenue
F-33 State Revenue
F-33 Federal Revenue
F-33 Total Expenditures
F-33 Current Instructional Programs Expenditures
F-33 Instruction Expenditures
F-33 Current NonInstructional Program Expenditures

Listing of NCES Working Papers to Date

Please contact Ruth R. Harris at (202) 219-1831
if you are interested in any of the following papers

Number	Title	Contact
94-01 (July)	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02 (July)	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03 (July)	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04 (July)	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05 (July)	Cost-of-Education Differentials Across the States	William Fowler
94-06 (July)	Six Papers on Teachers from the 1990-91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
94-07 (Nov.)	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
95-01 (Jan.)	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95-02 (Jan.)	QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95-03 (Jan.)	Schools and Staffing Survey: 1990-91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95-04 (Jan.)	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95-05 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
95-06 (Jan.)	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95-07 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95-08 (Feb.)	CCD Adjustment to the 1990-91 SASS: A Comparison of Estimates	Dan Kasprzyk
95-09 (Feb.)	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95-10 (Feb.)	The Results of the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95-11 (Mar.)	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95-12 (Mar.)	Rural Education Data User's Guide	Samuel Peng
95-13 (Mar.)	Assessing Students with Disabilities and Limited English Proficiency	James Houser
95-14 (Mar.)	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95-15 (Apr.)	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95-16 (Apr.)	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95-17 (May)	Estimates of Expenditures for Private K-12 Schools	Stephen Broughman
95-18 (Nov.)	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96-01 (Jan.)	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-02 (Feb.)	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96-03 (Feb.)	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
96-04 (Feb.)	Census Mapping Project/School District Data Book	Tai Phan
96-05 (Feb.)	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96-06 (Mar.)	The Schools and Staffing Survey (SASS) for 1998-99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96-07 (Mar.)	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96-08 (Apr.)	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96-09 (Apr.)	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998-99 SASS	Dan Kasprzyk
96-10 (Apr.)	1998-99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96-11 (June)	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96-12 (June)	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96-13 (June)	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96-14 (June)	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-15 (June)	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96-16 (June)	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96-17 (July)	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
96-18 (Aug.)	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
96-19 (Oct.)	Assessment and Analysis of School-Level Expenditures	William Fowler
96-20 (Oct.)	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96-21 (Oct.)	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96-22 (Oct.)	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96-23 (Oct.)	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96-24 (Oct.)	National Assessments of Teacher Quality	Dan Kasprzyk
96-25 (Oct.)	Measures of Inservice Professional Development: Suggested Items for the 1998-1999 Schools and Staffing Survey	Dan Kasprzyk
96-26 (Nov.)	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96-27 (Nov.)	Intersurvey Consistency in NCES Private School Surveys for 1993-94	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-28 (Nov.)	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
96-29 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96-30 (Dec.)	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-01 (Feb.)	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97-02 (Feb.)	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-03 (Feb.)	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97-04 (Feb.)	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-05 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-06 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-07 (Mar.)	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97-08 (Mar.)	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-09 (Apr.)	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97-10 (Apr.)	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993-94 School Year	Dan Kasprzyk
97-11 (Apr.)	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97-12 (Apr.)	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97-13 (Apr.)	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
97-14 (Apr.)	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97-15 (May)	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97-16 (May)	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97-17 (May)	International Education Expenditure Comparability Study. Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
97-18 (June)	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97-19 (June)	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97-20 (June)	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97-21 (June)	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
97-22 (July)	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-23 (July)	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97-24 (Aug.)	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97-25 (Aug.)	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97-26 (Oct.)	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
97-27 (Oct.)	Pilot Test of IPEDS Finance Survey	Peter Stowe
97-28 (Oct.)	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97-29 (Oct.)	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman
97-30 (Oct.)	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Steven Gorman
97-31 (Oct.)	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Steven Gorman
97-32 (Oct.)	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires)	Steven Gorman
97-33 (Oct.)	Adult Literacy: An International Perspective	Marilyn Binkley
97-34 (Oct.)	Comparison of Estimates from the 1993 National Household Education Survey	Kathryn Chandler
97-35 (Oct.)	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
97-36 (Oct.)	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-37 (Nov.)	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Steven Gorman
97-38 (Nov.)	Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey	Kathryn Chandler
97-39 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey	Kathryn Chandler
97-40 (Nov.)	Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey	Kathryn Chandler
97-41 (Dec.)	Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association	Steve Kaufman
97-42 (Jan. 1998)	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
97-43 (Dec.)	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
97-44 (Dec.)	Development of a SASS 1993-94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98-01 (Jan.)	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98-02 (Jan.)	Response Variance in the 1993-94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
98-03 (Feb.)	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98-04 (Feb.)	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
98-05 (Mar.)	SASS Documentation: 1993-94 SASS Student Sampling Problems; Solutions for Determining the Numerators for the SASS Private School (3B) Second-Stage Factors	Steven Kaufman
98-06 (May)	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98-07 (May)	Decennial Census School District Project Planning Report	Tai Phan

DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA